

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 30, 2002, 13:33:19 : Search time 206.87 Seconds
(without alignments)
11461.589 Million cell updates/sec

Title: US-09-997-610-1

Perfect score: 1381

Sequence: 1 gatagtgcatctactctctc.....tgtaacctcatgtatgtag 1381

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 1736436 segs, 858457221 residues 3472872

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802:*

1: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	665.4	48.2	2590	AAH14327	Human cDNA sequenc
2	642.2	46.5	3093	AAS03687	Rhesus gene locus:
3	642.2	46.5	9236	AAS03689	Rhesus gene locus:
4	642.2	46.5	9236	AAS03690	Rhesus gene locus:
5	642.2	46.5	9241	AAS03688	Rhesus gene locus:
6	621.4	45.0	6063	AAL03635	Human reproductive
7	620.4	44.9	3203	AAD05134	Human secreted pro
8	619.8	44.9	6063	AAL03634	Human reproductive
9	619.6	44.9	465237	ABA90193	Human oestrogen re

10	605.2	43.8	1736	AAH13678	Human cDNA sequenc
11	598	43.3	22509	AAS44505	Human LECTI DNA C1
12	598	43.3	31529	AAS44506	Human LECTI DNA C1
13	592.4	42.9	32127	AAI99255	Human excretory re
14	592.4	42.9	32127	AAI99255	Human kidney relat
15	592.4	42.9	72215	AAK86832	Human immune/haema
16	590.8	42.8	1278	AAAD05178	Human secreted pro
17	582.8	42.2	1485	AAS19984	DNA encoding novel
18	564.6	40.9	8387	ABAI4892	Human nervous syst
19	564.6	40.9	8387	ABAI4892	Human immune/haema
20	560.6	40.6	1580	AAS87262	DNA encoding novel
21	559.2	40.5	7726	ABAI21079	Human nervous syst
22	559.2	40.5	1621	AAS67924	DNA encoding novel
23	558.2	40.4	81369	AAA97997	Human T gene DNA.
24	551.6	39.9	1306	AAS64637	DNA encoding novel
25	551.4	39.9	3049	AAH16637	Human cDNA sequenc
26	545.4	39.5	1405	AAH70265	Human immune/haema
27	542.2	39.3	1405	AAK70266	Human immune/haema
28	525.4	38.0	1946	ABAI4227	Human breast cell
29	525.4	38.0	1946	ABAI4227	Human foetal liver
30	525.4	38.0	1946	ABA24459	Probe #2925 for ge
31	525.4	38.0	1946	AAK02964	Human brain expres
32	525.4	38.0	1946	AAK28410	Human bone marrow
33	525.4	38.0	1946	AAI12975	Probe #2908 for ge
34	525.4	38.0	1946	AAI34334	Probe #3020 used f
35	525.4	38.0	1946	AAI02893	Probe #2884 used f
36	521.8	37.8	22756	AAI03921	DNA encoding human
37	521.8	37.8	22756	AAI03921	Human reproductive
38	517.6	37.5	923	AAS66623	DNA encoding novel
39	509.8	36.9	26410	AAK70623	Human immune/haema
40	507.2	36.7	57728	AAK87588	Human 9p11 chromos
41	503.4	36.5	50000	AAD26400	Human glutamate re
42	503.4	36.5	50000	AAD26437	Human GRM3 gene fr
43	501	36.3	32986	AAK69758	Human immune/haema
44	501	36.3	32986	AAK84629	Human immune/haema
45	487.6	35.3	2197	AAH18389	Human cDNA sequenc

ALIGNMENTS

AAH14327	1	AAH14327	standard; cDNA; 2590 BP.
AC	AAH14327;		
DT	26-JUN-2001	(first entry)	
DE	Human cDNA sequence SEQ ID NO:11697.		
DE	Human cDNA sequence SEQ ID NO:11697.		
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.		
XX	Homo sapiens.		
OS			
PN	EP1074617-A2.		
XX			
PD	07-FEB-2001.		
XX			
PF	28-JUL-2000; 2000EP-0116126.		
XX			
PR	29-JUL-1999; 99JP-0248036.		
PR	27-AUG-1999; 99JP-0300253.		
PR	11-JAN-2000; 2000JP-0118776.		
PR	02-MAY-2000; 2000JP-0183767.		
PR	09-JUN-2000; 2000JP-0241899.		
XX	(HELI-) HELIX RES INST.		
XX			
PI	Ota T, Isogai T, Nishikawa T, Hayashi K, Yamamoto J;		
XX	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;		
DR	WPI: 2001-318749/34.		

PI Flegel WA, Wagner FF;

XX WPI: 2001-291052/30.

XX New nucleic acid molecular structure, useful for detection of common
PT RHD positive haplotypes in D-negative individuals, comprises RHD, SMP1
PT and RHE genes -

XX Example 10; Fig 5; 135bp; English.

XX The sequence represents the coding sequence of Rhesus gene locus:

CC RHD gene deletion in Rh negative haplotypes. The Rhesus genes

CC locus comprises the RHD, SMP1 and RHE (all undefined) genes and/or the

CC Rhesus box(es), preferably the hybrid Rhesus box, the upstream Rhesus box

CC and/or the downstream Rhesus box. The RHD and RHE genes are located at

CC chromosome 1 p34.1-p36. Rhesus box flanks the breakpoint region of the

CC RHD deletion in the common RHD negative haplotypes. The sequence has

CC been used to design primers which are useful for: (1) the specific

CC detection of the common RHD positive haplotypes in D-negative

CC individuals; (2) blood group typing; (3) determining whether a patient

CC can be transfused with RhD negative blood and whether blood is suitable

CC for transfusion to patients who should not be exposed to antigen C; (4)

CC assessing the risk of a RhD negative mother of conceiving or carrying an

CC RhD positive foetus. Anti-D antibodies are useful for treating pregnant

CC women who are Rhesus D negative, where the foetus is not homozygous for

CC the RHD gene to treat or prevent haemolytic disease of the newborn.

XX Sequence 3093 BP; 891 A; 754 C; 619 G; 829 T; 0 other;

Query Match 46.5%; Score 642.2; DB 22; Length 3093;

Best Local Similarity 79.1%; Pred. No. 2.4e-190;

Matches 844; Conservative 0; Mismatches 178; Indels 45; Gaps 5;

QY 352 gaatatttgctcaagaaggaagcaaatlttgcttaataagaagaatttcaagcaga 411
DB 1377 GAACCTTGAACCTTCAGAGAGATGATTAGATATCGAGAGAGAAATTTCAAGAGCA 1318
QY 412 aagcatcaagaagtgacttgcttgcttaagaagcatcagttcacaagaagcaga 471
DB 1317 AAGCATTCAGAGAGTGACTTGCTGCTTAAAGACATTCAGTTTATTAAGGAGACAGA 1258
QY 472 gataaagatltcagaanaatttgcacctacacatgtgataaagaagaanaa---ccattt 529
DB 1257 GCATTAAGATTCGAAATTTGACGCTGACATGATGATGAGAAAGAAATTTCCATT 1198
QY 530 tctgaggggaaatcaagcttgctcagaagaatttcataatgaatgaagagctgaatgt 589
DB 1197 TCTGAGGAGAAATTCAGCTGCTGACAGAAATTTGCATGAGTAA-CAGAGCCAAATGCT 1139
QY 590 aatcctcaagaagaatgaggaanaatctctcctgacatgtcagaagtlctacagaagctcc 649
DB 1138 AATTCCTCAAGCAATGAGGGAATATGCTCCAGGCAATGTCAGAGGCTTTATGCAACCC 1079
QY 650 atcaaatcactgctgagagcct-aggagaanaatggttttttgagacagagccagagtc 708
DB 1078 CTCCCATCACAAGGTCAGAGGATATCAGGAGAAATGTTTGTGGCCAGGCCGGGTC 1019
QY 709 cctgctgctgtgacagcagcagacttggtccctgtgtcccaagta----- 755
DB 1018 CTCATGCTGTGTGACGCCCTAGGAGACTTGTCCTGATCCAGCCACATCCCAACCATCA 959
QY 756 -----atcacgtgtgtcctcagaaggtgcaagccccaagcctt 794
DB 958 CTGACGGAGGAGCAAGTATGAGCTTGCGCTTGCGGAGTGCAACCCCAAGCCCTT 899
QY 795 ggcacatcccaagtgctgttgagcctgtgtgcaagaagtcgaagaatgtgagttggg 854
DB 898 GACACCTTCACAGTGTGTGTGAGACTGGAGTGCACAGAAAGCAAGAAATTTGGGTTTGA 839
QY 855 aaccccaatcagatcctcagaatatataagaaccctcagagtcaccaagagtttg 914
DB 838 AACCTTCCGCTTGAATTAAGAGAGATGTGCGGAAATGCTTGATGCCCAAGATTTG 779

QY 915 ctgtaagggtggtgtcctcatatgagaacccctcgaaggtaglacaaaaaggaaatgttg 974
DB 778 CTGACAGAGACAGGGCCCTCATGAGATCCTGTGCGAGGCGAGTGCAGAGGAAATGTGG 719
QY 975 ggtgtgagagcccccacacagaatgccagtggtgtcctcatctgtatgagtctgagaaga 1034
DB 718 GGTCAAGACACCCACACACAGTCCCTACTGGGGCACACCTAGTGAAGTGTGAGAAAGG 659
QY 1035 gtccacacatccctcagagatcccaagaaggtagatccactacagcctgtgcagatgtgc 1094
DB 658 G-----TCCTCCAGACCCCAAGATGATGATCCACGACAGCTTGACACCTGTACCTG 606
QY 1095 aaaaatccacagacactcagtgccagcctgtgaaagcagcagagatgagtgatccct 1154
DB 605 GAAAGCTTCACACACTCAATGCAGCCCATGAAGACAGCTGAGAGGAGGCTGTACCT 546
QY 1155 aaaaaacgtatgtgagagctgacccaagacgttggaatctacactctgtgactgtgcaag 1214
DB 545 GCAAGAGTACAGGGGCGAGGCTGCCACACACATGGGAAACCCCTTCATCTGCGTG 486
QY 1215 acctgagctgagacatgagatcaaaaagatcatttgagcttaaatgactgtgc 1274
DB 485 ACCTGATGTGAGATGTGAGATGCAAGAGATCATTTTGGAGCTTAAAGATTGACTGCC 426
QY 1275 ccactggaattcgactatattgaggcccgtaacccttgttlttgccaatttttccat 1334
DB 425 CCACGTGATTTGAGACTCTCATGCGGCTGTAGCCCTCTTGTGTTGACCAATTTATCCCAT 366
QY 1335 ttggaactgcgattattaccgaatgcctgttaccatctgtatgtg 1381
DB 365 TTGGAATGCTGTATTTACCAATGCTGTACCCCATTTGATTCATG 319

RESULT 3
ID AAS03689 standard; DNA; 9236 BP.
XX AAS03689;
XX 29-AUG-2001 (first entry)
DE Rhesus gene locus: upstream Rhesus box of D-positives.
XX Rhesus box: RHD positive; sequence length polymorphism: SSP; RHD; SMP1;
KW RHE; Rh negative; blood group typing; blood transfusion; antigen C;
KW haemolytic disease of the newborn; chromosome 1 p34.1-p36; ds.
XX Homo sapiens.
XX WO200132702-A2.
XX 10-MAY-2001.
XX 31-OCT-2000; 2000WO-EP10745.
XX 02-NOV-1999; 99EP-0121686.
XX 31-MAY-2000; 2000EP-0111696.
XX (DRKB-) DRK BLUTSPENDEDIENST BADEN WÜRTTEMBERG.
XX Flegel WA, Wagner FF;
XX WPI: 2001-291052/30.
XX New nucleic acid molecular structure, useful for detection of common
PT RHD positive haplotypes in D-negative individuals, comprises RHD, SMP1
PT and RHE genes -
XX Disclosure; Fig 9; 135bp; English.
XX The sequence represents the coding sequence of Rhesus gene locus:
CC upstream Rhesus box of D positives. The Rhesus genes locus

CC comprises the RHD, SMP1 and RHCE (all undefined) genes and/or the
 CC Rhesus box(es), preferably the hybrid Rhesus box. The upstream Rhesus box
 CC and/or the downstream Rhesus box. The RHD and RHCE genes are located at
 CC chromosome 1 p34.1-p36. Rhesus box flanks the breakpoint region of the
 CC RHD deletion in the common RHD negative haplotypes. The sequence has
 CC been used to design primers which are useful for: (1) the specific
 CC detection of the common RHD positive haplotypes in D-negative
 CC individuals; (2) blood group typing; (3) determining whether a patient
 CC can be transfused with Rhd negative blood and whether blood is suitable
 CC for transfusion to patients who should not be exposed to antigen C; (4)
 CC assessing the risk of a Rhd negative mother of conceiving or carrying an
 CC RHD positive foetus. Anti-D antibodies are useful for treating pregnant
 CC women who are Rhesus D negative, where the foetus is not homozygous for
 CC the RHD gene to treat or prevent haemolytic disease of the newborn.

XX Sequence 9236 BP; 2467 A; 2319 C; 2000 G; 2450 T; 0 other;

Query Match 46.5%; Score 642.2; DR 22; Length 9236;
 Best Local Similarity 79.1%; Pred. No. 4.5e-190;
 Matches 844; Conservative 0; Mismatches 178; Indels 45; Gaps 5;

QY 352 gaatttggcctaaggaagcaaatatttgctaaggaagaaatttctaagcaga 411
 DB 6053 GAACCTTGAACCTTCAGAGAGATGATTGAGTATCTGCGAGAGAAATTTCTAAGCAGA 5994
 QY 412 aagcattcaagagtgacttggtgtttaaagcattcaattcaaggaagcaga 471
 DB 5993 AACGATTCAAGAGGTGACTTGCTGTGTTAAAGCATTCAGTTTATTAAGGAGACAGA 5934
 QY 472 gcaatagattcaagaatttgcaacctgacatgataaataaagaaaaa--ccattt 529
 DB 5533 GCATTAAAGTTGGAAAATTTCACCTGACATGATGATGAAAAGAAAATTCCTATT 5874
 QY 530 tctgaggggaaattcaagctgctcagaatttgcatatgaatgaaggagctgaatt 589
 DB 5873 TCTGAGGAGAAATTCAGCTGCTGCTCAGAAATTTGCATGATGA--CAGAGCCAAATGCT 5815
 QY 590 aatctccaagaacaatggggaaaataatctcctgacatgltcagaagcttcacaagatcc 649
 DB 5814 AATTCCCAAGACAAATGGGAAAATGTCTCCAGGCAATGTCAGAGGCTTTATGCAACCC 5755
 QY 650 ataatactgctgctgagagcct-aggagaagaatttttgggagagccagaggttc 708
 DB 5754 CTCCTCATCAGAGCTCCAGAGGATATGAGAAAATGTTTGTGGCAGGCCCCGGGGTTC 5695
 QY 709 cctgtgctgtgtgcagccttaagaaactgtgctgtgtccagtta----- 755
 DB 5694 CTCATCTGTGTGACAGGCTAGGAGACTTGCTCCTCGATCCAGCCACATCCCAACATGA 5635
 QY 756 -----attcaactgtgtcttaagaagggttgaagcccaagcctt 794
 DB 5634 CTGACGGGAGGCAAGTAGAGCTTGCGCTGTAGCTTGGGGAGGTGCAAGCCCAACCTT 5575
 QY 795 ggcagcttccaagtgtgtgagcctgtggttgcagaagatcaagaattgaagtttgg 854
 DB 5574 GACAGCTTCCATGTGTGTGTGAGACTGCGAATGCACAGAAATTAATTTGGGGTTTGA 5515
 QY 855 aacctcaatcagatttcaagaatatatggaacccctgagatgcccagcagcaagtttg 914
 DB 5514 AACCTTGGCTAGATTAAAGAGAGATGCGGAATTCCTGAGATGCCAGTCAAGATTG 5455
 QY 915 ctgtaggggtgtgtgtcctcaatgagaaactctgcagaaggtagatcaaaagggaaatgttg 974
 DB 5454 CTTCAGAGGACAGGCGCTCATATGAGATCTCTGCAAGGCGAGTGAAGGGAATATGTG 5395
 QY 975 ggtgaggagccccaacagagttcccccgttgggtcatttaagtaagctgtgagaaga 1034
 DB 5394 GGTCAAGAGACCCACACAGAGTCCCTACTGTGGGACACCACTAATGAGCTGTGAGAGAG 5335
 QY 1035 gtccacatctcccaagatcacaagaggtatatacctacgtcagcttgagcgtgtg 1094
 DB 5334 G-----TCTTCACAGACCCCAAGATGTGTAGATCACCGACAGCTTGACCGTGTG 5282

QY 1095 aaaaatccacagacactcaagtccagccttgaagaacagcaggaatgagctgtaccct 1154
 DB 5281 GAAAAGCTGCAGACACTCAATGCCAGGCCATGAAGACAGCTGAGAGGGGGGTGACCT 5222
 QY 1155 aaaaacagctgtgtgcagagctgcagacacgcgtggaattcaactcttgcattgtcatg 1214
 DB 5221 GCAAAAGTACAGGGGCAAGCTGCCCCAAGACATGGGAAACCCATCCCTTCATGCTGTG 5162
 QY 1215 acctgagctgagacataagagtcacaagacatatttgaagcttaagattgactgccc 1274
 DB 5161 ACCTGAGATGTGAGATGTGGAATCAAGAGATCATTTTGGACTTTAAGATTGACTGCC 5102
 QY 1275 ccactgatttcgagactatataatggggccgtlacccttgttttggccaatttttccat 1334
 DB 5101 CCAGTGAATTTTGGACCTCATGAGGCTGTAGCCTCTTGTGTTTGAACCAATTAATTCAT 5042

QY 1335 ttgagacgctgattatcccaatgacctgaactcatgtatgtag 1381
 DB 5041 TTGGAATGGCTGTATTATTCACCAATGCTGTACCCCATTTGATCTAG 4995

RESULT 4
 AAS03690/C
 ID AAS03690 standard; DNA; 9236 BP.
 XX
 AC AAS03690;
 XX
 DT 29-AUG-2001 (first entry)
 XX
 DE Rhesus gene locus: downstream Rhesus box of D-positives.
 XX
 XX Rhesus box; RHD positive; sequence length polymorphism; SSP: RHD; SMP1;
 KW RHCE; Rh negative; blood group typing; blood transfusion; antigen C;
 KW haemolytic disease of the newborn; chromosome 1 p34.1-p36; ds.
 XX
 OS Homo sapiens.
 XX
 PN W0200132702-A2.
 PD
 PD 10-MAY-2001.
 XX
 PF 31-OCT-2000; 2000WO-EPI0745.
 XX
 PR 02-NOV-1999; 99EP-0121686.
 PR 31-MAY-2000; 2000EP-0111696.
 XX
 PA (DRKB-) DRK BLUTSPENDEDIENST BADEN WÜRTTEMBERG.
 XX
 PI Flegel WA, Wagner FF;
 XX
 DR WPI; 2001-291052/30.
 XX
 PT New nucleic acid molecular structure, useful for detection of common
 PT RHD positive haplotypes in D-negative individuals, comprises RHD, SMP1
 PT and RHCE genes -
 PS
 PS Disclosure; Fig 10; 135pp; English.

The sequence represents the coding sequence of Rhesus gene locus:
 CC downstream Rhesus box of D positives. The Rhesus genes locus
 CC comprises the RHD, SMP1 and RHCE (all undefined) genes and/or the
 CC Rhesus box(es), preferably the hybrid Rhesus box. The upstream Rhesus box
 CC and/or the downstream Rhesus box. The RHD and RHCE genes are located at
 CC chromosome 1 p34.1-p36. Rhesus box flanks the breakpoint region of the
 CC RHD deletion in the common RHD negative haplotypes. The sequence has
 CC been used to design primers which are useful for: (1) the specific
 CC detection of the common RHD positive haplotypes in D-negative
 CC individuals; (2) blood group typing; (3) determining whether a patient
 CC can be transfused with Rhd negative blood and whether blood is suitable
 CC for transfusion to patients who should not be exposed to antigen C; (4)
 CC assessing the risk of a Rhd negative mother of conceiving or carrying an
 CC RHD positive foetus. Anti-D antibodies are useful for treating pregnant

CC Women who are Rhesus D negative, where the foetus is not homozygous for
 CC the RHD gene to treat or prevent haemolytic disease of the newborn.
 XX
 XX Sequence 9236 BP; 2467 A; 2319 C; 2000 G; 2450 T; 0 other:

Query Match 46.5%; Score 642.2; DB 22; Length 9236;
 Best Local Similarity 79.1%; Pred. No. 4,5e-190;
 Matches 844; Conservative 0; Mismatches 178; Indels 45; Gaps 5;

```

OY 352 gaatatttgcgaagaagaagaatttggctaaataagaagaagaatttcaagcaga 411
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6053 GAACCTTGAACTTCAGAGAGATTTAGAGTATCTGCGAGAGAAATTTCTACAGCA 5994
OY 412 aagcaatcaagaagtgacttggtgctgttaagaagcatcaatttcaataagaagcaga 471
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5993 AAGCATTCAGAGGTGACTTGGTGCTGTTAAAGACATTCAGTTTATTAAGGAAAGCACA 5934
OY 472 gataaaggttcaagaagaatttgcacccctgaacaattgtataaagaaga---cccat 529
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5933 GCATAAAGTTTCGAAATTTTCACGCTGCACATGTGATGAGAAAGAAATTTCCATTT 5874
OY 530 tcttgagggaattcaagctggctgcagaatttgcataatgaatgaagagctgaatgt 589
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5873 TCTGAGGAGAAATTCAGCTGCTGCAGAAATTTGCATGAGTAA-CAGAGCCAAATGCT 5815
OY 590 aatcctcaagaacaatggggaataatactccttgacatgtcaagaagctcacaagatcc 649
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5814 AATTCACCAAGACAAATGGGGAATAATGCTCCAGGCAATGCAAGGCTTTATGCGCAACC 5755
OY 650 atcaaatcaactgagcttgagagcct-agaagaataatgttttcttgaggacagccaggtc 708
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5754 CTCCCATACAGAGTCCAGGTCATCAGAGAAATAATGTTTGTGGCCAGGCCGGGCTC 5695
OY 709 cctgtgctgtgtgagagccttaagaagcttgctgctgtgtcccaagta----- 755
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5694 CTGATGCTGTGTGACGCTAGGAGACTGTGCTGCCCTGTCATCCAGCCACTCCCAACCATGA 5635
OY 756 -----atcagctgtggtctcagaaggttgcaagagcccaagcctt 794
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5634 CTGACGGGAGGCAGAGTGAAGAGCTTGCGGTGAGCTTCGGGGAGTGCAAGCCCAAGCCCTT 5575
OY 795 ggcagcttccaaagtgtgttgagcctgtggtgcaagaagatcagaagattgagttggg 854
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5574 GACAGCTTCCAGTGTGTGAGACTGCGAGTGCACAGAAAGTCAAGAAATTTGGGGTTTGCA 5515
OY 855 aaccctcaatcagaattcagaagaatataatgaacccctggatgcccagaaggaagtttg 914
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5514 AACCTTCCGCTTGAATTAAGAGAGATGTGCGGAATGCTTGATGCCCATCTCAAGAAAGTTTG 5455
OY 915 ctgtaggggtgggtgctcctcaatgagaacctctgcagaaggttaataaagaaggaatgttg 974
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5454 CTGCGAGGAGAGGGCCCTCATGTGAGATCTCTGCGCAGGGCAATGCGAAGAAATGTGG 5395
OY 975 ggttgaggagccccaacaacagagttcccaagtgggtccatctagttagagctgtgagaaga 1034
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5394 GGTGAGAGACCCACACACAGTCCCTACTGGGGCACACCTGTGTGAGCTGTGCAAGAG 5335
OY 1035 gtccacacatcctcagaagctcagaaggttaataatcagaagcttgagagctgtggccg 1094
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5334 G-----TCTCCAGAGACCCCGAATGATGATCCACGACGAGCTTGACCCGTGTACCTG 5282
OY 1095 aaaaaatcacaagacactcagtgccagcctgtgaagaagcagaaggtatgagcttgaacct 1154
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5281 GAAATGCTGACACACTCAATGCTCAGCCCATGAAGCAAGCTAGAGGGAGGCTGTATACCT 5222
OY 1155 acaaaacgtagtgagcaagctgacccaagaacggtgggaatactactcctctgcatctcat 1214
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5221 GCAAAGGTACAGGGGCGACAGCTGCCCAAGACCATGGGAACCCACCTTCATCTGCGCTG 5162
OY 1215 accctgagcgtgagacatgaggttcaaaagatcatcttggagcctttaaagttaactgccc 1274
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5161 ACCTGATGTGATGTGAGTGTGAGTCAAGAGATCATTTTGGAGCTTTAAGATTGTACTGCC 5102

```

```

OY 1275 ccactgagttcgcgaactatagggcccgtaacccttgttttggccaatlttccat 1334
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5101 CCACGTGATTTGGACTCTCATGCGCCTGTAGCCCTTGTGTTGTGACCAATTTATCCCAT 5042
OY 1335 ttggaactgcgctattaccgaatgcctgtactcatcatgtagtg 1381
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5041 TTGGAAATGCTGTATTACCAATGCTGTATACCCCATTTGTATCTAG 4995

```

RESULT 5
 AAS03688/c
 ID AAS03688 standard; DNA; 9241 BP.

XX AAS03688;
 AC 29-AUG-2001 (first entry)
 XX
 DE Rhesus gene locus: hybrid Rhesus box of RHD negatives.

XX Rhesus box; RHD positive; sequence length polymorphism; SSP; RHD; SMP1;
 KW RHCE; Rh negative; blood group typing; blood transfusion; antigen C;
 KW haemolytic disease of the newborn; chromosome 1 p34.1 p36; ds.

XX Homo sapiens.
 OS
 PN WO200132702-A2.

XX 10-MAY-2001.

XX 31-OCT-2000; 2000MO-EP10745.

XX 02-NOV-1999; 99EP-0121686.

XX 31-MAY-2000; 2000EP-0111696.

XX (DRKB-) DRK BLUTSPENDEDIENST BADEN WÜRTTEMBERG.

XX Flegel WA, Wagner FF;

XX WPI; 2001-291052/30.

PT New nucleic acid molecular structure, useful for detection of common
 PT RHD positive haplotypes in D-negative individuals, comprises RHD, SMP1
 PT and RHCE genes -
 XX
 PS Disclosure: Fig 8; 135pp; English.

CC The sequence represents the coding sequence of Rhesus gene locus:
 CC hybrid Rhesus box of RHD negatives. The Rhesus genes locus
 CC comprises the RHD, SMP1 and RHCE (all undefined) genes and/or the
 CC Rhesus boxes), preferably the hybrid Rhesus box, the upstream Rhesus box
 CC and/or the downstream Rhesus box. The RHD and RHCE genes are located at
 CC chromosome 1 p34.1 p36. Rhesus box flanks the breakpoint region of the
 CC RHD deletion in the common RHD negative haplotypes. The sequence has
 CC been used to design primers which are useful for: (1) the specific
 CC detection of the common RHD positive haplotypes in D-negative
 CC individuals; (2) blood group typing; (3) determining whether a patient
 CC can be transfused with Rhd negative blood and whether blood is suitable
 CC for transfusion to patients who should not be exposed to antigen C; (4)
 CC assessing the risk of a Rhd negative mother of conceiving or carrying an
 CC Rhd positive foetus. Anti-D antibodies are useful for treating pregnant
 CC women who are Rhesus D negative, where the foetus is not homozygous for
 CC the RHD gene to treat or prevent haemolytic disease of the newborn.

SO Sequence 9241 BP; 2454 A; 2320 C; 2014 G; 2453 T; 0 other:

Query Match 46.5%; Score 642.2; DB 22; Length 9241;
 Best Local Similarity 79.1%; Pred. No. 4,5e-190;
 Matches 844; Conservative 0; Mismatches 178; Indels 45; Gaps 5;

```

OY 352 gaatatttgcgaagaagaagaatttggctaaataagaagaagaatttcaagcaga 411
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 6053 GAACCTTGAACTTCAGAGAGATGATTTAGAGTACTGGCCAGAGAAATTTCTAAACACCA 5994
Qy 412 aagcatlcaagaggtgacttggtgctgtltaagagcatlcaatgataagggaagcaga 471
Db 5993 AAGCATTCAGAGGAGTGGTGGTGTAAAGACATCAGTTTTATTAAGGAAGACGA 5934
Qy 472 gcaataaggtlcaagaataattgcacccctgacaatgcatataaagaagaaa--cccatlt 529
Db 5933 GCATTAAGATTGGGAAATTTGACCTGACATGTCATGTAAGAAAGAAATTCACATT 5874
Qy 530 tctgaggggaaattcaagctgctgcagaatitgcatatgataaggaagctgatatlt 589
Db 5873 TCTGAGGAGAAATTCAGCTGGCTGCAGAAATTTGCATGAGTAA -CAGAGCCCAATGCT 5815
Qy 590 aatcctcaagaacaatggygaaatatctcctgacatgltcaagaagctctcacagatgcc 649
Db 5814 AATTCACAGACAAATGGGAAATGTCTCCAGGGCAGTGCAGAGTCTTTATGGCAACCC 5755
Qy 650 atcaatcaatgctgctggaagcct-aagagaaaaatggttttctggaacagcccaaggtc 708
Db 5754 CTCCCATCACAGGTCACAGGATATCAGGAAAAATGCTTTGTTGGCCAGGCCGGGTC 5695
Qy 709 cctgtgtgtgtgagccttagagactggtgcccctgtgtccagta----- 755
Db 5694 CTCATGCTGTGTGACGCTTAGGACTGTGTGCTTCGATCCGACCCATCCCAACCATGA 5635
Qy 756 -----atcagctgtgcttcaagaagggtgcaagcccaagcctt 794
Db 5634 CTGACGAGAGCGAAGTATGAGACTTGGGCTGTAGCTTGGGGAGTGCACGCCCAAGCCTT 5575
Qy 795 gggagcttccaagtgtgtttgaagctgtggttgaaagaaggtcaaatatagagtttgg 854
Db 5574 GAGAGCTTCAGATGTGTGTGAGACTGTGAGTGCAGAGATCAAGAAATGGGTTTGGAA 5515
Qy 855 aaacctcaatcagatltcagaagaatalatgaaacccctcagatgacagccaaagagtttg 914
Db 5514 AACCTTCGCTGATTAAGAGAGATGTGCGGAAATGCTTGATGCCACTGCAGAGTTTG 5455
Qy 915 ctgtagaggtgaggtgctcctcatgagaaacctctgcgaaggtatgacaaagggaaatgttg 974
Db 5454 CTGCAAGACAGGAGGCTTCATGAGATCTCTGCGCAGGGCAGTGCAGAGGGAAATGTGG 5395
Qy 975 ggtgaggagcccccacagaagatccccaagtgggtcctcatatgaagctgtggaagaa 1034
Db 5394 GGTCAAGACCCACACACAGTCCCTACTGSGGCACTCACTAGTGAGCTGTGAGAGAG 5335
Qy 1035 gtccacatcctccagactcagagaaggtatagatccactgacaactgagcatgtgacctg 1094
Db 5334 G-----TCCTCAGACCCCAAGAAATGTGATCACCAGACACTTGACACGCTGACCTG 5282
Qy 1095 aaaaatccacagacactcaagtgcagcctgtgaaagcagcaggatgagctgtacacct 1154
Db 5281 GAAAGACTGCAGACACTCAATGATCCAGCCCATGAAGAGAGCTGAGAGGGAGCTGACCT 5222
Qy 1155 acaaaacgttagtgcaagctgacacaagacgctggaatcctaccccttgcatgtcatg 1214
Db 5221 GCAAAAGTACAGGGGCAAGCTGCCAAGACACAGGAGAACCCACCTTCATCTGCGTG 5162
Qy 1215 acctgagctgagacatgagtcacaaagagatcatitttgagctttaagatttgatgcc 1274
Db 5161 ACCTGATGTGAGATGTGAGATCAAGAGAGATCTTTTGGAGCTTTTAAATTTGACTGCC 5102
Qy 1275 ccaatgatttcgactatatatgggccctfcaaccctlttttgccaatttttccat 1334
Db 5101 CCACGTGATTTTGGACTCTCAATGGGCGTGAAGCCTGTTTGTTTTGAACAATTTATCCCAT 5042
Qy 1335 ttggaactgcgtatttaccacaagcctgtactcctcatgttatgtag 1381
Db 5041 TTGGAATGCTGTATTTACCAATGCCGTGACCCCAATTTATCTATCTAG 4995

RESULT 6
AAL03635

ID AAL03635 standard; DNA: 6063 BP.
XX
AC AAL03635;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 6323.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
OS Homo sapiens.
PN WO200155320-A2.
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001MO-US01339.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0203515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 01-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.

PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234224.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235833.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249219.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.

PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251858.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
PI WPI; 2001-465570/50.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition -
XX
PS Disclosure; SEQ ID NO 6323; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
SQ Sequence 6063 BP; 1655 A; 1162 C; 1259 G; 1987 T; 0 other:

Query Match 45.0%; Score 621.4; DB 22; Length 6063;
Best Local Similarity 79.9%; Pred. NO. 1.2e-183;
Matches 821; Conservative 0; Mismatches 166; Indels 40; Gaps 6;

QY 392 gaagaatttctaagcagaagaacatcctaagagtgacttggtgtgttaagcattc 451
|||||
DB 3481 gaagaatttctaagcagaagaacatcctaagagtgacttggtgtgttaagcattc 3540
QY 452 agttcataaggaggaagcagaataagatcagaataattgcacctgacaatgata 511
|||||
DB 3541 catttaaaaggga--aaagcataaaaaatttgnaaaatttcagcgcattatgagca 3598
QY 512 aaaaagaanaaccatttcgtgagggnaattcaagctgtcgcagaaattgcatagt 571
|||||
DB 3599 gaagagagaagacctgttctttagaggaattcaagctgtcgcagaaattgcataagt 3658
QY 572 aatgagagctgaatgtaattactcaagcaattgggnaaatctccgcagatctcag 631
|||||
DB 3659 tacagggaagctgaattgtaattccaagacaattgggnaaatctccgcagatctcag 3718
QY 632 aggtctcacagcagctcacaatcacaatcagctgtgagggcctagag--aaatggtt 688
|||||
DB 3719 aggtctcacagcagcctccatcacagaccctggaagactcagaggaagaaatggtt 3778
QY 689 tgtgggaagagcccaagggtcctgtgtgtgtgagcagtagaagattgtgctgtgtc 748
|||||
DB 3779 tgtgggcaagcccaagggtcctccatgtgtgtgtgagcagtagaagattgtgctgtgtc 3838
QY 749 ccaagttaattcag-----ctgggcttaagag 776
|||||
DB 3839 ccagctgtccagcattgtcctaagggcgaggtacacgtctccgctgggttctcagaag 3898
QY 777 gtcaagcccaagcctgtgcagcttcacagtggtgttagcctgtggtgcagaagaagt 836
|||||
DB 3899 gtcaagcccaagcctgtgcagcttcacagtggtgttagcctgtggtgcagaagaagt 3958
QY 837 caagaattgaggttgggaacctccaatcagaattcagaagaatatggaacccctgga 896
|||||
DB 3959 caagaattgaggttgggaacctccaatcagaattcagaagaatatggaacccctgga 4018
QY 897 tgcacagcagaagttgctgtgaggggtgtgtcctcattagaggaacccctggaaggtag 956
|||||
DB 4019 tgcacagcagaagttgctgtgaggggtgtgtcctcattagaggaacccctggtagggag 4078

[illegible]

RESULT 7

AAD05134
ID AAD05134 standard; cDNA; 3203 BP.

AC AAD05134;

DT 17-JUL-2001 (first entry)

DE Human secreted protein-encoding gene 14 cDNA clone H1SBF60, SEQ ID NO:24.

KW Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; hematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammatory; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiodysplasia; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder;
KW endocrine disorder; infection; wound healing; vulnery;
KW cell culture; chemotaxis; food additive; gene therapy;
KW blinding partner identification; ss.

05 Homo sapiens.

FH	Key	Location/Qualifiers

FT /product= "Human secreted protein"

PN WO200134769-A2.

PD 17-MAY-2001.

PF 01-NOV-2000; 2000WO-US30040.

XX 05-NOV-1999; 99US-0163580.
PR 30-JUN-2000; 2000US-0215130.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Ruben SM, Komatsoulis GA, Wei P, Baker KP, Fiscella M;
XX
XX WPI; 2001-308781/32.
DR P-PSDB; AAE01245.
XX
XX New isolated nucleic acid molecule encoding a human secreted protein is
PT used in preventing, treating or ameliorating a medical condition -
XX
XX Claim 1; Page 408-409; 519pp; English.
XX

AAD05121-AAD05203 represent cDNAs corresponding to 24 human secreted protein genes, and AAE01232-AAE0311 represent the proteins they encoded (CC AAE01312-AAE01340 represent human secreted protein variants or fragments). The secreted proteins and their genes are useful for preventing, treating or ameliorating medical conditions, e.g., by protein or gene therapy. Pathological conditions can be diagnosed by determining the amount of the new protein in a sample or by determining the presence of mutations in the new genes. Specific uses are described for each of the 24 genes, based on the tissues in which they are most highly expressed, and include developing products for the diagnosis or treatment of proliferative disorders, cancer, tumours, foetal and developmental abnormalities, haematopoietic disorders, diseases of the immune system, AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation, allergies, neurological disorders (e.g., Alzheimer's disease, Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders, angioneurotic disorders, kidney disorders, gastrointestinal disorders, pregnancy-related disorders, endocrine disorders, and infections. The proteins can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues, to identify their cognate ligands or binding partners, and in chemotaxis, and can be used as a food additive or preservative to modify storage properties. Antibodies specific for a protein of the invention can be used in alleviating symptoms associated with the disorders mentioned above, and in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The present sequence represents a human secreted protein-encoding cDNA of the invention.

Sequence 3203 BP; 868 A; 650 C; 740 G; 943 T; 2 other;

Query Match	44.9%;	Score 620.4;	DB 22;	Length 3203;
Best Local Similarity	78.3%;	Pred. No. 1.7e-183;		
Matches 835; Conservative	0;	Mismatches 192;	Indels 40;	Gaps 6

Qy	352	gaatatttggcctaagaggaagcaaatatttgctataataagaagaanaatttctaagaaga	411
Db	1599	gaactttgactcttgagagaagatgttttaagggtatcttgcgtcgaagaattcttaagaaga	1655
Qy	412	aagcattcaagagggtgacttggctgtcttaagaagcattcagttctcaaaaggagaagcaga	471
Db	1659	aagcattcaagagggtgatttgggtgctgttaagaagcattcattttaaaggga--aaca	1715
Qy	472	gcataaggtctagaanaatttgcacccctgacatgttgataaaaaagaanaaccatttc	531
Db	1717	gcataaanaatttgganaaatttgcagccagctbatbtaacgacaagaaggaaaacctgtttt	1776
Qy	532	tgaagggtgaattcaagctgctctgaagaatttgcatagtatgaaagaagctgaatttaa	591
Db	1777	tgaagggtgaattcaagctgctcgtagaanaatttgcataagtttaacgggaagctgaatttaa	1833
Qy	592	tccctcaagacaattggganaaatatctctcttgacatgtctcaagagctcttcaagaagctcat	651
Db	1837	tccctcaagacaattggganaaatgtctctccagggcatgtctcaagagctcttcaagaagccct	1896

[illegible][illegible]

Db 4278 tgaagcccaaggggccagagctgcgcccaagactaaagggaaccctactccttgcattcaattgta 4337
Oy 1216 cctgaagctggaacatgagatcgaacaaagatacattcttgagacttaagattgacgcc 1275
Db 4338 cctgagatgtggaacatgagatcgaagagagatcatcttgcgaagcttaataattgaccgcct 4397
Oy 1276 cactgattctgcacctataatbgyggcccgct-acccctctgttcttgccaattttccat 1334
Db 4398 cgcctgattctgacacttgcatbgyggccttgtaagcccttctgttgyggcaacttccocat 4457
Oy 1335 ttggaactgcgcgtatttaccacaagctctgtaacctccatttatagtag 1381
Db 4458 ttggaatgctgctgatttaccatcactgtaaccctatgtaag 4504

RESULT 9
ABA90193
ID ABA90193 standard; DNA: 465237 BP.
XX
XX ABA90193:
XX AC
XX DT 11-FEB-2002 (first entry)
XX
DE Human oestrogen receptor alpha gene.
XX
XX Human; oestrogen receptor alpha; ESR-alpha; ER; chromosome 6; Syc-2;
KW synaptic nuclei expressed gene 2; haplotype; cytosolic; osteopathic;
KW cardiant; vasotrophic; gene therapy; vaccine; cancer; osteoporosis;
KW cardiovascular disease; oestrogen receptor; ds.
XX
XX Homo sapiens.
OS
PN WC200162969-A2.
XX
PD 30-AUG-2001.
XX
XX 20-FEB-2001; 2001WO-US05358.
XX PF
XX 22-FEB-2000; 2000US-183756P.
PR 20-OCT-2000; 2000US-0692414.
PR 24-JAN-2001; 2001US-0768184.
XX
PA (PEKE) PE CORP NY.
XX
XX Kalush F, Cassel MJ, Hwang SS, Wlun-Deen ES;
XX
XX WPI: 2002-041152/05.
DR P-PSDB; AAG68251.
XX
XX Novel variant of estrogen receptor alpha polypeptide useful for
PT determining the biological activity of a protein for high throughput
PT screening and for raising antibodies that elicit an immune response in
PT host -
XX
XX Example; Fig 1 page 1-93; 333pp; English.

The present invention describes an isolated peptide (I) consisting of an amino acid sequence selected from: (a) the amino acid sequence of a variant of the oestrogen receptor alpha (ESR-alpha) protein in AAG68251; or (b) a fragment comprising at least 10 contiguous amino acids of the protein in AAG68251. (I) has cytosolic, osteopathic, cardiant and vasotrophic activities, and can be used in gene therapy and vaccine production. (I) is useful for identifying an agent that binds to (I), by contacting (I) with an agent and assaying the contacted mixture to determine whether a complex is formed with the agent bound to the peptide. A polynucleotide (II), encoding (I), is useful in the development of diagnostics and therapies for diseases and disorders mediated/modulated by an oestrogen receptor (ER). (II) is also useful in gene therapy for treating cancer, osteoporosis and cardiovascular diseases. The human ESR-alpha gene is located on chromosome 6. The present sequence represents the human ESR-alpha gene, which is given in the exemplification of the present invention.

||||| ||||||| ||||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 439687 gcttagaattgactgcgcgtgtgatttagactgtgtggtccctgtaaccccttg 439746
QY 1315 tttagccatttttccatttggaactgcgcgtattaccatgcttactcatgtg 1374
Db 439747 tttagccatttttccatttggaactgcgcgtattaccatgcttactcatgtg 439806
QY 1375 tatgta 1380
||| ||
Db 439807 tatcta 439812

RESULT 10
AAH13678
ID AAH13678 standard; cDNA; 1736 BP.
XX
AC AAH13678;
XX
XX 26-JUN-2001 (first entry)
DY
XX
DE Human cDNA sequence SEQ ID NO:10539.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000BP-0116126.
XX
PR 29-JUL-1999; 993P-0248036.
PR 27-AUG-1999; 993P-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakematsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
DR
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX Claim 8; SEQ ID 10539; 2537bp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.
XX
SQ Sequence 1736 BP; 449 A; 372 C; 453 G; 462 T; 0 other;

Query Match 43.8%; Score 605.2; DB 22; Length 1736;
Best Local Similarity 79.1%; Pred. No. 6,8e-179;
Matches 813; Conservative 0; Mismatches 168; Indels 47; Gaps 6;

QY 392 gaagaattcttaagcagaagcattcaagagtgactgtgtgtgttaagagcttc 451
|||||
Db 111 gaagaattcttaagcagaagcattcaagagtgactgtgtgtgttaagagcttc 170
QY 452 agttcataagggagcagagcctaaagagttcagaatattgcacctgacaatgtgata 511
|||||
Db 171 agttcataagggagcagagcctaaagagttcagaatattgcacctgacaatgtgata 230
QY 512 aaaaagaataaaccatttcttgagggaattcaagctgtgtgtgtgcaaatgtgata 571
|||||
Db 231 gacaagaataaaccatttcttgagggaattcaagctgtgtgtgtgcaaatgtgata 290
QY 572 aatgagagcgtgaattgttaattcctcaagcaatgagggaataatctcttgagactgac 631
|||||
Db 291 agcaagagcgtgaattgttaattcctcaagcaatgagggaataatctcttgagactgac 350
QY 632 aggtctcacagcagctccatcaataatcactgtgtgtgtgtgtgtgtgtgtgtgt 691
|||||
Db 351 agacctcacagcagctccatcaataatcactgtgtgtgtgtgtgtgtgtgtgtgt 405
QY 692 gggacagcccaaggttcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 751
|||||
Db 406 gggcctgt 465
QY 752 gtaattc-----agctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 775
|||||
Db 466 gctgt 525
QY 776 ggtgtcaagcccaagcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 835
|||||
Db 526 ggtgtcaagcccaagcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 585
QY 836 tc-aaggaattgaggt 894
|||||
Db 586 tcaaggaattgaggt 642
QY 895 gatgccagcagaagattgt 954
|||||
Db 643 gatgccagcagaagattgt 701
QY 955 agtacaaaagggaattgt 1014
|||||
Db 702 agtgtgaagggaattgt 761
QY 1015 tagtagactgtgtgaagaagagtcacatcttcagagctcagaagaggtagactatga 1074
|||||
Db 762 tagtagactgtgtgaagaagagtcacatcttcagagctcagaagaggtagactatga 821
QY 1075 cagcttgagaatgt 1134
|||||
Db 822 cagcttgagaatgt 881
QY 1135 aggaatgagctgt 1194
|||||
Db 882 cggagggagagctgt 941
QY 1195 ctactcttgatgt 1254
|||||
Db 942 tcaacttgatgt 1001
QY 1255 agcttaagatttgaagcccaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1313
|||||
Db 1002 agcttaacatttgaagcccaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1061

Oy	1314	gctttggcccaatttcttcatttggaactgcgcgtattatccaacagcctgtaacctcatt	1373
Db	1062	gctttggcccaatttcttcatttggaactgcgcgtattatccaacagcctgtaacctcatt	1121
Oy	1374	gtaatgtag_1381 	
Db	1122	gtaatgtag_1129	
RESULT 11			
ID	AAS44505	standard; DNA: 22509 BP.	
AC	AAS44505;		
DT	18-DEC-2001	(first entry)	
XX			
DE	Human LEKTI DNA clone C1T978SKB_94F21 contig 11, SPINK5 exons 1-4.		
KW	Human; SPINK5; lympho-epithelial Kazal-type related inhibitor; LEKTI; ds;		
KM	serine protease inhibitor; atopic disease; Netherton's syndrome; asthma;		
KW	eczema; hayfever; antisthmatic; anti-allergic; anti-inflammatory;		
KW	dermatological; PCR primer; sequencing primer; gene therapy.		
XX			
OS	Homo sapiens.		
PN	WO200164747-A1.		
PD	07-SEP-2001.		
PF	02-MAR-2001; 2001MO-GB00897.		
PR	02-MAR-2000; 2000GB-0005098.		
PR	03-MAR-2000; 2000GB-0005229.		
XX			
PA	(ISIS-) ISIS INNOVATION LTD.		
PI	Hovnanian A, Chavanas S, Cookson W, Moffat M, Walley A;		
DR	WPI: 2001-582149/65.		
XX			
PT	Determining susceptibility to atopic disease or carrier status of		
PT	Netherton's syndrome in humans by identifying variants of or mutations		
PT	in SPINK5, a gene encoding lympho-epithelial Kazal-type related		
PT	inhibitor -		
XX			
PS	Disclosure: Page 88-94; 123pp; English.		
XX			
CC	Sequences AAS44359-AAS44514 represent the SPINK5 gene, contigs and		
CC	fragments of a SPINK5 clone, sequencing primers and PCR primers for		
CC	SPINK5. SPINK5 encodes lympho-epithelial Kazal-type related inhibitor		
CC	(LEKTI), a serine protease inhibitor. Susceptibility or predisposition to		
CC	an atopic disease in a human subject can be detected by screening the		
CC	genome for one or more polymorphic variants of SPINK5 gene and/or		
CC	expression of a variant LEKTI protein in a tissue. Carrier status of a		
CC	subject or development of Netherton's syndrome is diagnosed by screening		
CC	for the presence of loss-of-function mutations in the SPINK5 gene. An		
CC	expression vector comprising a nucleic acid encoding a serine protease		
CC	inhibitor or its functional fragment can be used in screening for		
CC	compounds with potential pharmacological activity by determining the		
CC	serine protease activity of a protein previously identified as a ligand		
CC	of the LEKTI protein. The atopic diseases include Netherton's Syndrome,		
CC	asthma, eczema and hayfever.		
XX			
SO	Sequence 22509 BP; 6725 A; 4291 C; 4284 G; 7209 T; 0 other;		
Query Match	43.3%; Score 598; DB 22; Length 22509;		
Best Local Similarity	77.9%; Pred. No. 5.6e-176;		
Matches 815; Conservative 0; Mismatches 210; Indels 21; Gaps 77			
Oy	gaataatctgctaactgagaagcaaatcttggtcataaagagaagaatcttaacagca	411	

Db	14148	gaactttaacttcagagagatgatattaaaggtatcctgtagcagaagaattcttaagcagca	14207
QY	412	aagcactcaagggctgactgtggtctgtcttaagagcatccaattccaataaggagagcagca	471
Db	14208	ggcattctcaaggagataaactgtgtgtctcttaagagcatctcagctttttaaggagagcagca	14267
QY	472	gcataagattcagaanaatttgcacctgtacaaatgtatataaaaaaga--aaaccattt	529
Db	14268	acataaagattgtgaaaatttgcagctgcagctgcagctgtgataaaaaaagaataaccattt	14327
QY	530	tcctgggggaaattccaagtgtcgcagaaatttcgatactatgataagagagctgtatgtt	589
Db	14328	tcctggaggagaaattcaaggccagtgtcataaatttgcataaagttagcaaggagccgtgatt	14387
QY	590	aatctccaagacaatctggagaaatattctcctgtacatgtgcagaggtctccaagcagctcc	649
Db	14388	aattccccaagacaat--gggaaaatgtctccaagagatatcatcagagaccttctgtgcagccc	14446
QY	650	atcaaatcaactgtgacctgtgagcgtcctagag--aaatgtgttctgtgggacagagccaggg	706
Db	14447	ctcccatcattacatgtcccaagaggttagagagaaaaaatgtctctcctaaagccagcttaagg	14506
QY	707	tcctctgtctgtgtgtcagcc-----tagagacttgtgtgccctgtgtccagttaa	756
Db	14507	tcctctctgtctgtgtgcagccctgtcgttgaagacactccaagtgtgcgaagggggcatatgac	14566
QY	757	ttcagctctgtggttcagaagggtgcagaagccccaagccttgcagagcttccaagtggtctga	816
Db	14567	ctgggctatgtgtcttcgaagggtgcagaagccttgaagccttgcagagcctccaatgtgtgtga	14626
QY	817	gctctgtgtgtgcagaagaagtcacaagaatttgaagttctgggaaacctccaatcagaattcca	876
Db	14627	gctctgtgtgtgcagaagaagtcagaagcttgaagaaacctccaagctcagctagttcca---	14683
QY	877	gataatgtgaaacccctgtgatagtccagagcagaagtttctgtatgaggtgtgggtctctcatg	936
Db	14684	gataatgtgaaatgcctgtgatttccaagcagaagaatttctgtatgaggtgtgggtctctcatg	14743
QY	937	ggagaacctctgtgaagggtgtatcaaaaggaaatttgggtgtggggagccccaacagaagt	996
Db	14744	ggagaacctctgtgaagggtgtatcaaaaggaaatgttgggtgtggggagccccaacagaagt	14803
QY	997	ccccagctggggcttccatctagttagagctgtgtgaaagaagtcacaacatctctcagaactcca	1056
Db	14804	ccctact--gggacacactgtgtgtgagctgtgtgaagaagagagccttcaactcctcagaacca	14862
QY	1057	gaagagtgatccatccatgtcagagctgtgcagatgtgcctgtaaaaaaccacagacactcagtg	1116
Db	14863	gaatagcagatccacaggtaggtttgcacatatagtcctgtagaagaagccacaacactcagtg	14922
QY	1117	ccaagctcttgaaagcagcagaggtatgtagttctgtacccctaaanaaccgtatgtgtgcgaact	1176
Db	14923	ccaagctcttgaaagcagcctgtgagagagcggttaaccctgtcaaaagccacaagggtgtaggtt	14982
QY	1177	gaccaagacgtgtggaatctacactcttgcattgtcatgtcatcactgtgcagctgtgaagatvgag	1236
Db	14983	gaccaagacgtgtggaatcccaacctcttgcacacagatgactcgtgatagtagaacataagatt	15042
QY	1237	caaaagagatcaattttgagctttaagaatttgcactgtgcccaactgtgatttcgactatatt	1296
Db	15043	caaaatatacatattttggaactttaagttttgaactgtccttcotgatatttcgactgtgat	15102
QY	1297	gggggcccgtta--ccccttggttttggccaatttttccatttggaaatcgcgtatttaacc	1355
Db	15103	ggcactcttaagcccttgtgttctggccaattctcccaatttggaaacagctgtatattgaccc	15162
QY	1356	aatgccttacctcatgtatgtag 1381	
Db	15163	aatgctgttaccctcatgtatctag 15188	

ID AAS44506 standard; DNA; 31529 BP.
XX AAS44506;
AC AAS44506;
XX 18-DEC-2001 (first entry)
DT
XX
XX Human LEKTI DNA clone C17978SKB_94F21 contig 8, SPINK5 exon 5.
DE
XX
XX Human; SPINK5; lympho-epithelial Kazal-type related inhibitor; LEKTI; ds,
KW serine protease inhibitor; atopic disease; Netherton's syndrome; asthma;
KM eczema; hayfever; antiasthmatic; antiallergic; antiinflammatory;
KM dermatological; PCR primer; sequencing primer; gene therapy.
XX
OS Homo sapiens.
XX
XX WO200164747-A1.
XX
XX 07-SEP-2001.
XX
XX 02-MAR-2001; 2001WO-GB00897.
XX
XX 02-MAR-2000; 2000GB-0005098.
XX 03-MAR-2000; 2000GB-0005229.
XX
XX (ISIS-) ISIS INNOVATION LTD.
XX
XX Hovanian A, Chavanas S, Cookson W, Moffat M, Walley A;
XX WPI; 2001-582149/65.
XX
XX
XX Determining susceptibility to atopic disease or carrier status of
PT Netherton's syndrome in humans by identifying variants of or mutations
PT in SPINK5, a gene encoding lympho-epithelial Kazal-type related
PT inhibitor -
XX
XX
XX Disclosure; Page 95-103; 123pp; English.
XX
XX Sequences AAS44359-AAS44514 represent the SPINK5 gene, contigs and
CC fragments of a SPINK5 clone, sequencing primers and PCR primers for
CC SPINK5. SPINK5 encodes lympho-epithelial Kazal-type related inhibitor
CC (LEKTI), a serine protease inhibitor. Susceptibility or predisposition to
CC an atopic disease in a human subject can be detected by screening the
CC genome for one or more polymorphic variants of SPINK5 gene and/or
CC expression of a variant LEKTI protein in a tissue. Carrier status of a
CC subject or development of Netherton's syndrome is diagnosed by screening
CC for the presence of loss-of-function mutations in the SPINK5 gene. An
CC expression vector comprising a nucleic acid encoding a serine protease
CC inhibitor or its functional fragment can be used in screening for
CC compounds with potential pharmacological activity by determining the
CC serine protease activity of a protein previously identified as a ligand
CC of the LEKTI protein. The atopic diseases include Netherton's Syndrome,
CC asthma, eczema and hayfever.
XX
XX
XX Sequence 31529 BP; 9686 A; 5798 C; 5934 G; 10111 T; 0 other;
SQ

Query Match 43.3%; Score 598; DB 22; Length 31529;
Best Local Similarity 77.9%; Pred. No. 6,8e-176;
Matches 815; Conservative 0; Mismatches 210; Indels 21; Gaps 7;

Db 14328 tctgagagaaattcaagccagctgtcataaatttgcataagtagaagagcctgaagtt 14387
||||| ||||||| ||| ||||||| ||| |||||
QY 590 aatcccaagaacattggggaataatctctcgtgacatgtcagaaggtcttcaacagatcc 649
||||| ||||||| ||||||| ||||||| ||| |||||
Db 14388 aatcccaagaacatt-gggaaaaatgtctccaagagatatcagaaccttcttgagcagcc 14446
QY 650 atcaaatcactgtgcttgagagcctbaggag--aaaatgttttggagacagccagagg 706
Db 14447 ctcccatcactgtcccgaggttttaagagaaaaatgttctctcagccaggttaggg 14506
QY 707 tccctgtctgtgtgcagcc-----tagagactgtgtgcccgtgtgtccaggttaa 756
||||| ||||||| ||||||| ||| |||||
Db 14507 tccctcgtctgtgtgcagccctgtgtgaagaactccaagtgtgtcgaagggtgccaatggcc 14566
QY 757 ttcagctgtgtctcagaagggtcgaagcccaagccctgtgcagcttccaagtgtgtctga 816
|| ||| ||||||| ||||||| ||||||| |||||||
Db 14567 ttgggtctatgtgtcttcgaagggtcgaagccttgaaagccttcgaagccttcacatgtgtctga 14626
QY 817 gctctgtgtgtcgaagaagtcagaagattgaagtttgggaacctccaatcagattcagaa 876
Db 14627 gctctgtgtgtcagagaagtcagaagcctgaaggtttgaaaaactcagcttagatctca--- 14683
QY 877 gatatatgaaacccctggatgtcccaaggcagaagatttgcgtgaagggtgggttctctaag 936
Db 14684 gatatatgaaatgtccgtgatgtccagagcagaagttctctgtaagggtgtgaggtccccaag 14743
QY 937 gagaacctctgcaagggtagtacaagaaggaaatgttgggttgaggccccacacagaagt 996
||||| ||||||| ||||||| ||| |||||
Db 14744 gagaacctctgtagggcagtagtacaagaagaatgttgagtagtgagccccacacagaagt 14803
QY 997 ccccaatgtgtgtccatctcgtatgagctgtgagaagaagtcacacatctctcagaactcca 1056
||| ||| ||||| ||| ||||||| |||||
Db 14804 cccatct-gggcacccactgtgtgtgagctgttagaagaagagcccttatctctccaagaacca 14862
QY 1057 gaaaggtagatccactcagcagctgtgcagcagctgtgccttgaaataatccaagaactcagtt 1116
||||| ||||||| ||||||| ||| |||||
Db 14863 gaatagcagatcccaaggttagctgtgacacatatgtcctgagaagccacaacactcagttg 14922
QY 1117 ccaagctgtgaaacaagaaggtgagttctgtacccatacaaaacgtagtgtgcagagct 1176
||||| ||||||| ||||||| ||| |||||
Db 14923 ccaagctgtgaaagcagctgtggagggagcgtgtacccctgcaagcagaaggtgtgaggttc 14982
QY 1177 gaccaagcgcgtgtgaaatctacactctgtcatgtatgacgtgtgagcgtgtgaaatctgagtt 1236
||||| ||||||| ||||||| ||| |||||
Db 14983 gcccaagccactgtgaaacccaactctgtcatcagcatgcagctgtgataagacataagatt 15042
QY 1237 caaaagagataattttagagcttaagatttgaactgtccccaatgatttggagattat 1296
||||| ||||||| ||||||| ||| |||||
Db 15043 caaaatataatcatttggaaacttgaagattgtgactgtccttcgaattctgcagctgtcat 15102
QY 1297 ggggcccgtta-ccccttgttttggccaatttttccatttggaaactgtcgtaattacc 1355
|| ||| ||||||| ||||||| ||| |||||
Db 15103 ggcacatgttagcccttgttttggccaatatctccatttggaaacagctgtatagccc 15162
QY 1356 aatgcctgtacctccatgtatgtag 1381
||||| ||||||| ||||||| ||| |||||
Db 15163 aatgcctgtacctccatgtatgtag 15188

RESULT 13
AA199255/c
ID AA199255 standard; DNA; 32127 BP.
XX
XX AA199255;
AC
XX 07-JAN-2002 (first entry)
XX
XX Human excretory related polynucleotide SEQ ID NO 1019.
DE
XX Human; nootropic; neuroprotective; cytotstatic; dermatological; virulide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antischlicking; antianaemic; antithrilitic; cancer;

CC (AA16291-AA163793) and the encoded polypeptides (AA42417-AA42691)
CC collectively known as kidney antigens and the use of such kidney antigens
CC for detecting disorders of the kidney, especially kidney cancer and
CC kidney cancer metastases. The polynucleotides and proteins are also
CC useful for preventing, treating or ameliorating medical conditions
CC e.g. by protein or gene therapy. The genes are isolated from a range
CC of human tissues disclosed in the specification. The nucleic acids,
CC proteins, antibodies and (ant)agonists are useful in the diagnosis,
CC treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer,
CC and other cancers of the adrenal gland, bone, bone marrow, breast,
CC gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders
CC e.g. Addison's disease, allergies, autoimmune haemolytic anaemia,
CC autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple
CC sclerosis, rheumatoid arthritis and ulcerative colitis;
CC (c) cardiovascular disorders such as myocardial ischaemias; (d) wound
CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC and (f) infectious diseases such as viral, bacterial, fungal and
CC parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 32127 BP: 8788 A: 6237 C: 6877 G: 10225 T: 0 other;

Query Match 42.9%; Score 592.4; DB 22; Length 32127;
Best Local Similarity 79.1%; Pred. No. 3.9e-174;
Matches 813; Conservative 0; Mismatches 161; Indels 54; Gaps 7;

QY 391 ggaagaatttcttaagcgcgaagaagatccaagaagtgaactggtgctgttaagcatt 450
DB 22042 GGAAGAAATTTCTAAGCAGACAGCATTCAGAGTACTTGGTGCTGTTAAAGGCATT 21983
QY 451 caattccataagggagagacataagatgcagaatctcagccctacacatgcat 510
DB 21982 CACTTTTAAAGGAAACAGACGACATAAAGCTTGGAAGTTTGACGCTGACATGCAAT 21923
QY 511 aaaaagaagaaaccattctctgaggggaaatccaagctggtcgcagaaattgtcatatg 570
DB 21922 AGAAAGAAAGAAAA-----TCTCAGGTGACATTCAGGACGCTGCAAGAAATTGGCATAG 21871
QY 571 taatggaggagctggaatgttaatcctcaagagaatgggaaataatcctccgagcatgta 630
DB 21870 TAATGGAGGGCCAAATGTTAATCCCAAGAAATGGGAAATGTCCTCCAGGGCATGTCA 21811
QY 631 gaagctctacagcagctccatcaatcaatcagctggtgagcctgaagag--aaatgctt 687
DB 21810 GAGATCTTCATGGCAGCCCTCCCATCAAGGCTTGCGAGGCTTAGAGGAAAGTGCTT 21751
QY 688 ttgtggagacagagcccaaggttccctgtctgtgtgacgctcagaagaactgtgctccgtgt 747
DB 21750 TCATGGGCGAGGCCAGGGTCCCATCTGCTGTGAAGCTTAGGGACTTGGCTCGCAT 21691
QY 748 ccaagtaattcaagc-----tgtgcttcaga 774
DB 21690 CTAGGCCGCTTAAGCCATGCGAGAAAGGGCCCAAGCTTAGAGCTCAGCATGTGGCTTCA 21631
QY 775 ggggtgcaagcccaagcctctgagcagctccaagtgtgtgagcctgtggtgcaagaa 834
DB 21630 GGGTGCAAGCCCAAGCCTTGCGACGCTT--CACATGAGTGTGAGCGCTTCAAGTCACTTA 21572
QY 835 gtcaagaattgaggttgggaacctcaatcagatctcagaagaatataagaaacctgt 894
DB 21571 GTCAAAGACTGGGGTTGGGAACCTCCACCTAGATTTCAAGAGTGTATGGAATCCCTG 21512
QY 895 gattcccaagcagaagattgtcgttaggggtggtgctcctcctgagaacctctcagaaggt 954
DB 21511 GATGCCCAAGCAGAA--TTTGCTGCTAGGGCAAGGGCCCTTAAGGAGACCTCTTAAGGGC 21453
QY 955 agtcaaaagggaatgctgtggtggaagcccccacacagagatcccccagtggtggtccatc 1014
DB 21452 AGTGTAAAGGAAAGTGTAGAGTGGTCCGCCACACAGAGTTCTCTATGTGACACTGTGCC 21393

QY 1015 tagtagagctgtgagaagaagtcacacatcctccagactccagaagggtagatcactga 1074
DB 21392 TAATGAGAGCTGTGAGAAAGAGAGCCAGCATCTCCAGATCCCAAGATGTAGATCCACTGA 21333
QY 1075 cagcttgagcagctgtgctggaataaatacagagacactcagctgcagcctgtgaagcagc 1134
DB 21332 CAGCTTGCACACGCTGCTGGAAAGAGCTGCAAGACATTCAGACACAGCTGTAAAGCAGC 21273
QY 1135 aggaatgagctgtgaacctcaaaaaccgtagtgcag--agctgacacaagacggtgggaa 1193
DB 21272 CAGGAGGAGAACTGTACCTCGACAGGCGACAGGGGTGAGAGCTGCCAGACATGGGAA 21213
QY 1194 tctacctgtgcatgtgcaatgacatgacgtgacgltgagacatgtagtcaaaagatcattg 1253
DB 21212 CCACCGCTTCATCAGCGTGAACCGGATGCCGATGAGACACAGGAATCAAGAGATCAATTTG 21153
QY 1254 gagcttgaagattgacgcgcacactgcatattggaattataggggcccgcgaacctt 1313
DB 21152 GACCTTTAAGATTTGACGTGCTCACTGATTTACACTT-----GCCTGTAGCCCTT 21100
QY 1314 gtttggcaatttttccatttgaaactgcgatttaccacatgcgttaccatcalt 1373
DB 21099 TGTTTAGCTAATTTCTCCATTTGGAAATGGCCGATTTTACCAATGCCGTACCCCAAT 21040
QY 1374 gtaatgag 1381
DB 21039 GTATCTAG 21032

RESULT 15
AAK86832/C
ID AAK86832 standard; DNA: 72215 BP.
XX
AC AAK86832;
XX
DE 07-NOV-2001 (first entry)
XX
DE Human Immune/haematopoietic antigen genomic sequence SEQ ID NO:41644.
XX
KW Human: Immune; haematopoietic; Immune/haematopoietic antigen; cancer;
XX
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PE 09-AUG-2001.
XX
PR 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232086.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235835.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.

PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 06-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Barash SC, Ruben SM;
PI
XX
XX
XX
DR WPI: 2001-483426/52.
XX
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
XX
PS Disclosure: SEQ ID NO 41644; 3071bp + Sequence Listing; English.
XX
XX
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK62170 to AAK91921. (I) have cytosolic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK62169
CC represent sequences used in the exemplification of the present invention.
XX
XX
SQ Sequence 72215 BP; 22256 A; 13595 C; 14340 G; 22024 T; 0 other;

Query Match 42.9%; Score 592.4; DB 22; Length 72215;
Best Local Similarity 79.1%; Pred. No. 6.4e-174;
Matches 813; Conservative 0; Mismatches 161; Indels 54; Gaps 7;

OY 391 ggaagaatttctaagcagaagcattcaagaagtgtggtgctgtttaagcatt 450
|||||
Db 22042 GGAAAGAAATTTCTAAAGACGACGATTCAGAGGTTGGTGTCTTTAAAGGCATT 21983
OY 451 cagtttcaagaaggagagacataaggttcagaanaattgcacccctgacacatgtgat 510
|||||
Db 21982 CAGTTTAAAGAAAGAAACAGACATATAAAGCTTGGGAAGTTTGCAGCCTGACATGCAT 21923
OY 511 aaaaaaagaaaccatttctctgagggaattcaagctgtgctcagaanaatttgcatalg 570
|||||
Db 21922 AGAAAAAGAAA-----TCTCAGGTGACATTCAAGGACGCTGCAGAAATTTCATAG 21871
OY 571 taatgaaggagctgaatttaattcaatcacaagacaattggggaataatctctgacatgtca 630
|||||
Db 21870 TAAATGAGGGGCCAAATGTTAATCCCAAGAAAATGGGAAAATGTTCTCCAGGGCATGTCA 21811
OY 631 gaggtcttcacagcagctccatcaatcaactggtcctgagagcctaaggag---aaatggtt 687
|||||
Db 21810 GAGATCTTCATGAGCCGCCCTCCATCACAGGCTTGCAAGGCTAGAGGAAAAAGTGTT 21751
OY 688 ttgttgagacagagcccaagggtccctgtgtgtgtgcaagacttggtgcccctgtgt 747
|||||
Db 21750 TCATGGGCGAGGCGCCAGGCTCCCATGCTGTGTGAAGCCTTAGGACTTGTGCTCCCTGCAT 21691
OY 748 cccaatttaattcagc-----tgtagcttcaaga 774
|||||
Db 21690 CTCAGCCGCTCTAGGCCATGCGCAGAAAAGGGCCCAAGCTAGAGCTCAGCATGTGCTTCAGA 21631
OY 775 ggggtcagaagccccaagccttggcagcttccaaagtgtgtgtgagcctgttggtcagaagaa 834
|||||
Db 21630 GGGTCAAGGCCCAAGCCTTGCGCAGCTT-CACATGTTGTTGAGCTTGCACAGTCA 21572
OY 835 gtcaagaattgaggttgggaacctccaatccagatttcagaagaatatatggaacccctgt 894
|||||
Db 21571 GTCAAGAACTGGGGTTTGGGAACCTCCACCTAGATTTCAGAAAGATGTAGAAATGCCCTG 21512
OY 895 gatgccagagcagaagttgtcgttaggggttggtcctcatgagaaacctgtcagaaggt 954
|||||
Db 21511 GATGCCACAGGCGAGAA-TTTGCTGCATGGGCAAGGCCCTTATGAGAACCTCTACTAGGGC 21453
OY 955 agtacaaaaagggaattgttgggttggagcccccacacagagtcgccagtggtggtctccalc 1014
|||||
Db 21452 AGTGTGAAAAAGGAAATGTGAGGTGCGGTGCCCCACACAGAGTTCTTACTGAGACATGCC 21393
OY 1015 tagtagaactgtgagaagaagtccacacatccctccagactccagaaggttagatccactga 1074
|||||
Db 21392 TAAATGAGCTGTGAGAGAGAGGCGCAGATCTCCAGATCCAGAAATGTTAGATCCACTGA 21333
OY 1075 cagcttgcagcatgtgcttgaanaaaatccacagacacactcagtgccagcctgttgaagcagc 1134
|||||
Db 21332 CAGCTTGCACACAGCTGTGGAAAGCTGAGACATTCACAGACAGCCTGTGAAGCAGC 21273
OY 1135 agggatgaggtcgtcaacctcaaaaaaccgtagtggcag-agctgaaccaagacgttggaa 1193
|||||
Db 21272 CAGGAGGGAAACTGTACCTCGAGCCACAGGGGTAGAGCTGCCCAAGACATGGGAA 21213
OY 1194 tctaacttctgcatgtcatgacccctggagcgtgagacatgaggtcaaaagaatcaatttg 1253
|||||
Db 21212 CCCACCGCTTGATCAGGTGACCGGATGCGAGACACGGAATCAAGAGATCATTTTG 21153
OY 1254 gagcttgaagatttgaactgcccccaatttcggaactatatatgsggcccgtaccccttt 1313
|||||
Db 21152 GAGCTTTAAGATTGACTGCTCCACTGGAATTCACACTT-----GCCGTGAGCCCTT 21100
OY 1314 gtttggccaatttttccatttgaactgacgatatlaaccaatgacctgtatccctcaatt 1373
|||||
Db 21099 TGTTAGCTAATTTCTCCATTGGAATGGCGGTATTTACCAATGCTGTACCCCATTT 21040

OY 1374 gtagttag 1381
|||||
Db 21039 GTATCTAG 21032

Search completed: June 30, 2002, 14:44:24
Job time: 4265 sec


```
Db 182 CTACTGGAAGATACAGAGAGAAATGTGGTTTGGAGCCCCCAACAGATATCCCTCTA 241
Qy 1005 gggctccatctagtagagctgtgagaagaagtcacacatccctccagactccagaagtg 1064
Db 242 GAACACTGCTATGAACTGTGAGAGAGATGGCCACTGCTATCCAGACACCGAATGATA 301
Qy 1065 gatccactgagagctgtgagcat-gtgcctgaaataatcccaacactgaagccagcc- 1122
Db 302 GACCCAGCAAAACTTATGCCATATGGCTATTAACCTACACACACTCAATGCCCAGCCC 361
Qy 1123 --tgtaaagcagcagggatgagtcgtctacccaacacgtaatggcagaagtcagc 1180
Db 362 CATTGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 421
Qy 1181 aagaacgtgtggaactacactctgtcatgtcaatgagcgtgagc-tgagacatggagtc 1239
Db 422 CAGGCCATGGAAGACAGACTCTTATATC AATGTACCTGATGATTTGAGACATGATCCN 481
Qy 1240 aagaagtcatttggagcttaagatc---tgactgccccactggaattcggaattata 1295
Db 482 AAGAAATCTTTTAACTCCACGGTTTAAATGACTGCCCTATTAATTCNGAATTAANA 541
Qy 1296 t-9gggccccgtaacccttgttttgccaattttccatttgaactgcgtatttacc 1354
Db 542 TCCNGGCGCTGTGACCTCTTTGCTTTGGCCATTCGCCCTTTTGGATGGCTMTTTTTC 601
Qy 1355 caatgcctgtaccctcat 1372
Db 602 CCATGCCCTGCCCTCTT 619

RESULT 5
US-09-128-155-16
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; EARLIER FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1) --(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16
```

Query Match 17.4%; Score 240.2; DB 3: Length 152331;

Best Local Similarity 59.4%; Pred. No. 1e-65; Matches 320; Indels 71; Gaps 7;

```
Qy 445 ggcattcaagttccatcaaggagcagagacagatgagctggaataattgacccctgaca 504
Db 71252 ggcacctatggttaaatggaagtagagctgacaaatttggaaatttcgacccctggc 71311
Qy 505 tgtgataaaaaaagaatacccatcttcgagggggaatccaagctgtgcgcagaataatgg 564
Db 71312 tttgtgcaagaagaataccaagtagcgtgcagagcaatcatgtcag---agagattg 71367
Qy 565 catatgtatgtggagcgtgaatgtaattccctcaagaataatggggaataataatctctgac 624
```

```
Db 71368 catgactaaaggagcccaagtgcttaattatcaagacaaagttaaaagccttgagggc 71427
Qy 625 atgtcagaaggtcttcacaagaagtcacaaatcacctggccttgagagcctag-----ga 677
Db 71428 atttcaagactctatgaagcagccctccatcacagctgacagaggtttgtgcactag 71487
Qy 678 gaaatgttttggggacagccaggtctcctgtgcgtgtgtgagactgagagacttg 727
Db 71488 ccagaagttcttaaggccanngccagggccacacgtgcataagcacttcgagacgtg 71547
Qy 738 tgcctgtgtccagtttaa----- 756
Db 71548 cgcgccgcatcccaagccacactcgtctcgtctccacccttgctcaaaagggccaagatag 71607
Qy 757 ---ttcagctgtgtgtcttcagaagggtgcgaagccccaagccttgagcacttccaagtgt 813
Db 71608 agcttggacccactgcctcccgagggcacaagcacaataagccttggtgttccatgt 71667
Qy 814 tgaagctgtgtgtgcaagaagaagtcagaat--tgaggtttgggaacctccaatcagatt 871
Db 71668 taagcttgcaggtgcacaagaatgcagaatlgagggagcttgcactccacctaatt 71727
Qy 872 cagaagacataatgaaacccctgagatgcccagcagaagcttgcctgtaggggtggtcc 931
Db 71728 cagagagatgtgtcagaacacctaggttccaggcagaaacatgatacaggggcagagccc 71787
Qy 932 tcaatgagaacctctgcagaaggtatgatacaaaagggaatgtgtgggtggagccccacac 991
Db 71788 ttgcagagaaacctctactaaggcaatgccaagaaatgtgtgggttgagatctctcacac 71847
Qy 992 aagatccccagtggtgtccatctctagtagagctgtgagaagaagtcacacatctccaga 1051
Db 71848 atgtgtcccatctggggaactacactctgtgatactgtggaaatggggtgtgtccctcaga 71907
Qy 1052 ctccagaaggttagatcacactgacagacttgcagcatgtcctgaaataccacagacat 1111
Db 71908 ccccaagatgtgtagatgactcgtgcagctgtgcacacctgacctggaaagctgcagcact 71967
Qy 1112 cagtgcagcctgtgaaagcagcaggtatgagtcgtacatccctacaacacgtgtgtgca 1171
Db 71968 caatcccaacccaagatgagctacgacacatgggtac--tccagggaaagccacagagca 72025
Qy 1172 gagctgcacaagaccgtgtgggaatctacactctgtcatgtcagacttgcagtgagatc 1231
Db 72026 gggcgtctcaaggccttggagcctacccctgt-----aacacgttgcagagcat 72076
Qy 1232 ggaatcacaagaagatcatcttggagctttaagattgactgcccacatggattcagact 1291
Db 72077 ggaatcaa---agattatgttgcagccttaagcttaagcttaatgatttccctgcaattcaa 72133
Qy 1292 tatatggggcccgtaacccttgttttggccaatttttccatttggaaactgcgtaatt 1351
Db 72134 cttgtgtggagcgtgtgttttttttttttttttttttttttttttttttttttttttt 72193
Qy 1352 acc 1354
Db 72194 acc 72196
```

RESULT 6

US-08-687-080-59

; Sequence 59, Application US/08687080

; Patent No. 5965427

; GENERAL INFORMATION:

; APPLICANT: Gregory Dolganov

; TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof

; NUMBER OF SEQUENCES: 175

; CORRESPONDENCE ADDRESS:

ADDRESSEE: Dehlinger & Associates

STREET: 350 Cambridge Avenue, Suite 250

CITY: Palo Alto

STATE: CA

COUNTRY: USA

Qy 599 gacaatgggaaataatctctggaacatgctcagaggtcttcacagacgtccatcaaatca 658
Db 1222 RRR 1163
Qy 659 ctgacctggagagccctagagaagaatggttggtgagcagcccgagg 706
Db 1162 RRR 1115

RESULT 8

US-08-463-911-6
; Sequence 6, Application US/08463911
; Patent No. 5869330
; GENERAL INFORMATION:
; APPLICANT: Scherer, Philipp E.
; APPLICANT: Lodish, Harvey F.
; TITLE OF INVENTION: A NOVEL SERUM PROTEIN PRODUCED
; TITLE OF INVENTION: EXCLUSIVELY IN ADIPOCYTES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Millita Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,911
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WHI95-05
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1313 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 73..804
; US-08-463-911-6

Query Match 3.5%; Score 48.4; DB 2; Length 1313;
Best Local Similarity 48.8%; Pred. No. 2.1e-05;
Matches 164; Conservative 0; Mismatches 166; Indels 6; Gaps 1;

Qy 42 atgtagaagtgtgctgagctccagcacacccccagagcccccagagaagatgaggtcctctg 101
Db 281 ATCCAGGTCTTATTGTCTTAAGGAGACATCGGTGAACCGAGATACCCGGGCGTGAAAG 340
Qy 102 gtgaccaggtttaccacaataatacagagaataaagtgaatgaacaaatgccccctgc 161
Db 341 GTCCCGAGAGCTTTCCTGGGAATCCAAAGCAGAGAAAGAACTGGAAGAGTGCT--- 397
Qy 162 ctgatatagaaggtcagccttactctggaagcctcagtgaacaaactcctctccttca 221
Db 398 ---ATGTATACCGCTCAGCATTCAGTGGGATGGAAGCTTACGTACTATCCCAACA 454
Qy 222 agcccatcatcttcacagaggttcctgtacaaatgccagagaggttaaaagagggcatg 281
||||| ||| || ||| ||||| ||| ||| |||

Db 455 TGCCCATTCGCTTACCAAGATCTTACATATCAGCAAAACACTGTGATGGCTCCACTG 514
Qy 282 gagctcttgcttcagagggtccttggaattactactcagcttgatgtgagctgcatc 341
Db 515 GTAATATCCATCGCAACATTCCTGGGCTGATACTTTCCTACATCACATCAAGCTTATA 574
Qy 342 attgcaaggtgaatatcttgcttaattgaggaacaa 377
Db 575 TGAAGATGTGAAGCTCAGCCTCTTCAAGAGACA 610

RESULT 9

US-09-140-804-9
; Sequence 9, Application US/09140804
; Patent No. 6197930
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Humes, Jacqueline M.
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
; FILE REFERENCE: 97-49
; CURRENT APPLICATION NUMBER: US/09/140,804
; CURRENT FILING DATE: 1998-08-26
; EARLIER APPLICATION NUMBER: 60/056,983
; EARLIER FILING DATE: 1997-08-26
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 4517
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-140-804-9

Query Match 3.5%; Score 48.4; DB 4; Length 4517;
Best Local Similarity 48.8%; Pred. No. 4.4e-05;
Matches 164; Conservative 0; Mismatches 166; Indels 6; Gaps 1;

Qy 42 atgtagaagtgtgctgagctccagcacacccccagagcccccagagaagatgaggtcctctg 101
Db 235 atccaggtcttatgtctccttaaggagacatcgttgaacccgagatcccggtgtagag 294
Qy 102 gtgaccaggtttaccacaataatacagagaataaagtgaatgaacaaatgccccctgc 161
Db 295 gtcccgaggtcttcgggaatccaagcagaagaagagagacctgtgaagaagtgcct--- 351
Qy 162 ctgatatagaaggtcagccttactctgtgaagctcagtgtgaacaaactcctccttca 221
Db 352 ---atgtataccgctcagcatcagtggtgaggtgagactcgttactatccccaaca 408
Qy 222 agcccatcatcttcacagaggtcctgtacaaatgccagagaggttaaaagagggccatg 281
Db 409 tgcctatgcttaccagaagctctctacatcagacaacacctaagtgtgctccatg 468
Qy 282 gagctcttgcttcagagggtccttggaattactactcagccttgatgttgcgtcatc 341
Db 469 gtaaatctcagcgaacatctctgtgctgtactactcttgcttaccatcaacagctata 528
Qy 342 attgcaaggtgaatatcttgcttaattgaggaacaa 377
Db 529 tgaagatgtgaaggtcagcctcttcaagaagaca 564
||||| ||| || ||| ||||| ||| ||| |||

RESULT 10

US-09-188-930-217
; Sequence 217, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murlison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells

;; TITLE OF INVENTION: and Methods for Their Use
;; FILE REFERENCE: 11000.1011c1
;; CURRENT APPLICATION NUMBER: US/09/188.930A
;; CURRENT FILING DATE: 1998-11-09
;; NUMBER OF SEQ ID NOS: 348
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 217
;; LENGTH: 1107
;; TYPE: DNA
;; ORGANISM: Rat
US-09-188-930-217

Query Match 3.1%; Score 42.4; DB 3; Length 1107;
Best Local Similarity 50.5%; Pred. No. 0.0016;
Matches 103; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

OY 169 agaaagtcacagcttactgtagaagctcagtcgagaaactcctctccttcaagcccat 228
DB 614 agccaagtcgagcctcttcgtagcgtaaccaagagttaccacgtagagcctgcccat 673
OY 229 catcttcaaggggctcctgtacaaatgcccagaagagattaaagagagccatggaagctt 288
DB 674 caagtttgaacaagattcgaatgaatgagagagagccactacatgcatccagtgagcaagt 733
OY 268 tgcctgcagagtgctgaggaattactaccagctttagttgagctgcatcattgcaaa 348
DB 734 cgtctgcagcgctgcagagatctacttactctatgacatgaacgtgccaacaaca 793
OY 349 ggtgaatatgtgctaagaagaa 372
DB 794 cctggcatcgccctcagtcgacaa 817

RESULT 11
US-08-463-911-1

;; Sequence 1, Application US/08463911
;; Patent No. 5869330

;; GENERAL INFORMATION:
;; APPLICANT: Scherer, Philipp E.
;; APPLICANT: Lodish, Harvey F.
;; TITLE OF INVENTION: A NOVEL SERUM PROTEIN PRODUCED
;; TITLE OF INVENTION: EXCLUSIVELY IN ADIPOCYTES
;; NUMBER OF SEQUENCES: 7
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
;; STREET: Two Millia Drive
;; CITY: Lexington
;; STATE: Massachusetts
;; COUNTRY: USA
;; ZIP: 02173

;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/463.911
;; FILING DATE:

;; CLASSIFICATION: 530
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Granahan, Patricia
;; REGISTRATION NUMBER: 32,227
;; REFERENCE/DOCKET NUMBER: WH195-05
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (617) 861-6240
;; TELEFAX: (617) 861-9540
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1276 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear

;; MOLECULE TYPE: DNA (genomic)
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 46..786
US-08-463-911-1

Query Match 2.9%; Score 40; DB 2; Length 1276;
Best Local Similarity 49.1%; Pred. No. 0.01;
Matches 106; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

OY 162 ctgatataaaggtcagcttctactgtagaagctcagtcgagaaactcctctccttca 221
DB 377 cttatattgacgttcacagcttcagtggtgagagaccgctcactgttccaatg 436
OY 222 agccatcctcttcaagaggtcctgtacaaatgcccagaagagattaaagagagccatg 281
DB 437 taccatttgcttttactgaatcttctacacacagatattttgacggagcagcagc 496
OY 282 gagctcttctgcagaggtgctggaattactaccagctttagttgagctgcatc 341
DB 497 gcaagttcttactgcaacatctccggacttacttcttaccacatcagcgtgtaca 556
OY 342 attgcaagtgaaattatgtgctaagaggaagcaaa 377
DB 557 tgaagatgtgaagtgacaccttctcaagaaagcaca 592

RESULT 12
US-08-543-246B-8

;; Sequence 8, Application US/08543246B
;; Patent No. 6262244

;; GENERAL INFORMATION:
;; APPLICANT:
;; TITLE OF INVENTION: DNA and amino acid sequence specific for
;; TITLE OF INVENTION: natural killer cells
;; NUMBER OF SEQUENCES: 24
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Michael W. Giynn
;; ADDRESSEE: No. 6262244artis Corporation
;; STREET: 564 Morris Avenue
;; CITY: Summit,
;; STATE: NJ
;; COUNTRY: US
;; ZIP: 07901-1027

;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/543.246B
;; FILING DATE:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/676,663
;; FILING DATE: 28-MAR-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US92/02469
;; FILING DATE: 27-MAR-1992

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/122,514
;; FILING DATE: 24-SEP-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Kassehoff, Melvyn M.
;; REGISTRATION NUMBER: 26,389
;; REFERENCE/DOCKET NUMBER: 118-7704/PCT/CONT
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 908-522-6927
;; TELEFAX: 908-522-6955
;; INFORMATION FOR SEQ ID NO: 8:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1755 base pairs
;; TYPE: nucleic acid

```
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 339..986
NAME/KEY: mat_peptide
LOCATION: 339..986
US-08-543-245B-8
```

```
Query Match 2.8%; Score 38.8; DB 4; Length 1755;
Best Local Similarity 62.2%; Pred. No. 0.03;
Matches 61; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
```

```
QY 1284 ttgcgaactatattggggccgcgtaccccttgcttggccaattttccatttggaactg 1343
DB 1564 ttgtgacttgcttgagtcgccatgactggttctcctccttcttcccttttggaatag 1623
QY 1344 ccgattaccgaatgcctgtaccccttgatgtatgtag 1381
DB 1624 taatatccatccatgattgttgcctccatgattgattttg 1661
```

```
RESULT 13
US-09-140-804-1
; Sequence 1, Application US/09140804
; Patent No. 6197930
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Humes, Jacqueline M.
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
; FILE REFERENCE: 97-49
; CURRENT APPLICATION NUMBER: US/09/140,804
; EARLIER FILING DATE: 1998-08-26
; EARLIER APPLICATION NUMBER: 60/056,983
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1347
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (198)...(926)
US-09-140-804-1
```

```
Query Match 2.7%; Score 37.4; DB 4; Length 1347;
Best Local Similarity 51.5%; Pred. No. 0.071;
Matches 86; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
```

```
QY 206 ctccctctcccttccaagccatcctccaagggtcctgtacaatgcccagagagat 265
DB 543 cctccgcgccttgacgaccccttgccctgcacgcgtgctgtgaaagagagagacat 602
QY 266 ttaagaagagcattggagcttctgttcgaggggtgcctggaattactactccactt 325
DB 603 taagacgcgcctacgcgaagttaccctgcagagtgcttggtggtctactactcgcgc 662
QY 326 gatgtgagctgcatcatgtcaagtgataattgctaatgagaa 372
DB 663 catgcacccgtctaccggcagcagcctgcatgtatctgtgtaagaa 709
```

```
RESULT 14
US-09-140-804-10
; Sequence 10, Application US/09140804
; Patent No. 6197930
```

```
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
APPLICANT: Humes, Jacqueline M.
TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
FILE REFERENCE: 97-49
CURRENT APPLICATION NUMBER: US/09/140,804
CURRENT FILING DATE: 1998-08-26
EARLIER APPLICATION NUMBER: 60/056,983
EARLIER FILING DATE: 1997-08-26
NUMBER OF SEQ ID NOS: 47
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 10
LENGTH: 729
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Degenerate nucleotide sequence encoding the zslg39
US-09-140-804-10
```

```
Query Match 2.6%; Score 36.2; DB 4; Length 729;
Best Local Similarity 27.9%; Pred. No. 0.12;
Matches 86; Conservative 33; Mismatches 189; Indels 0; Gaps 0;
```

```
QY 65 gcacacccagggcccccagaagaagtgaggccctctgctgacacaggttaccacatat 124
DB 205 ggmgncngngytnccngncmngngngaycngncmngngngagcngngncn 264
QY 125 acaggagaataagtgatgacaaatgcccctgctcgtatagaaagggcagcctt 184
DB 265 gcngnccnagcngcngcngngngatgysngtncnccmngnsgntlysgn 324
QY 185 actgtaagctcagtggaatacttccttccttccaagccatcattccaggggtc 244
DB 325 aarmngngarmngmngtncncnccmngngaycngncnctlygaymngtn 384
QY 245 cgtacaatgcccaagagatttaagagcattgagagcttctgtcaggggtgct 304
DB 385 yltgnaagacagcagcagcngngatgysngtncnccmngnsgntlygacngtn 444
QY 305 ggaattactactcagctgagctgagctgcatcatgcaatgcaagtgaaatgctga 364
DB 445 gngntataytatgtygngtncaygcnagcngtntaymngcnsmntncartcygaytn 504
QY 365 atgagaa 372
DB 505 glnaaraa 512
```

```
RESULT 15
US-08-627-151A-15
; Sequence 15, Application US/08627151A
; Patent No. 586341
; GENERAL INFORMATION:
; APPLICANT: SPINELLA, Dominic
; APPLICANT: BECHERER, Kathleen
; APPLICANT: BROWN, Steven
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; SCREENING DRUG LIBRARIES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Gen-Probe Incorporated
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/627,151A
; FILING DATE: 03-APR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fisher, Carlos A
; REGISTRATION NUMBER: 36,510
; REFERENCE/DOCKET NUMBER: CBI016
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-410-8926
; TELEFAX: 619-410-8928
; TELEX:
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1074 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-627-151A-15

```

```

Query Match          2.6%; Score 35.8; DB 2; Length 1074;
Best Local Similarity 54.1%; Pred. No. 0.2;
Matches 73; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 11 atacctgtcttaataacgcaagtcattgagcaatgtagaagttgctgagcctccagcac 70
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 768 ATATCGGGGTGAACGCTCAAGACATTCACACATGATGATGTCAGAGACCTCCAGCATCA 827

Qy 71 cccaggcccccaagaagtgaggccctcctggtgcaccaggttaccacatatacagga 130
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 828 CTGTGTCAATCCAGACGCTGAGCGGCGCTGAGGCACGCTGTGACGCTTGCCCGAGGA 887

Qy 131 gaaataagtaaatg 145
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 888 GGAGTTCGGGCAAGG 902

```

Search completed: June 30, 2002, 14:37:29
Job time: 6695 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 30, 2002, 11:48:13 ; Search time 1615.67 seconds
(without alignments)
11536.573 Million cell updates/sec

Title: US-09-997-610-1
Perfect score: 1381
Sequence: 1 gatagtgctacacgtctc.....tgacctccatgtagtag 1381

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estdb:*
2: em_esthum:*
3: em_estlin:*
4: em_estinu:*
5: em_estnu:*
6: em_estov:*
7: em_estro:*
8: em_estl:*
9: gb_estl:*
10: gb_estl2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_lin:*
15: em_gss_pln:*
16: em_gss_vitc:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	ID	Description
1	677.8	49.1	2615	BE420422	BE420422 32-393 hu
2	615.8	44.6	2009	BC004496	BC004496 Homo sapi
3	576.8	41.8	1514	BE512633	BE512633 32-1514 h
4	483	35.0	1036	BM471183	BM471183 AGENCOURT
5	479.2	34.7	870	BI488505	BI488505 603021014
6	477.8	34.6	898	BM457166	BM457166 AGENCOURT
7	462.6	33.5	1050	BM472108	BM472108 AGENCOURT
8	430.4	31.2	717	AG086951	AG086951 Pan trogl
9	419.8	30.4	676	AG064424	AG064424 Pan trogl
10	411.4	29.8	654	AG112666	AG112666 Pan trogl
11	411	29.8	721	BE728616	BE728616 601563827
12	407.8	29.5	935	BG681769	BG681769 602627877
13	405.4	29.4	760	BG281182	BG281182 602402107
14	404.6	29.3	736	BI754555	BI754555 603023384
15	404.2	29.3	691	AG062447	AG062447 Pan trogl
16	403.8	29.2	900	BM449699	BM449699 AGENCOURT
17	402.2	29.1	1182	BM451317	BM451317 AGENCOURT

18	401.6	29.1	729	10	BM045873	BM045873 603624875
19	401.2	29.1	701	12	AQ588873	AQ588873 CTRB1-E1-
20	401.2	29.1	704	12	AQ588828	AQ588828 CTRB1-E1-
21	398.4	28.8	720	12	AG116938	AG116938 Pan trogl
22	397.6	28.8	697	12	AG018926	AG018926 Homo sapi
23	394.4	28.6	699	9	AL602279	AL602279 DKFZP68B
24	394.2	28.5	833	9	AU140358	AU140358 AU140358
25	394	28.5	877	10	BI916691	BI916691 603178655
26	393.8	28.5	677	12	AG097258	AG097258 Pan trogl
27	392.4	28.4	659	12	AG092858	AG092858 Pan trogl
28	390.8	28.3	639	12	AG115932	AG115932 Pan trogl
29	389.4	28.2	676	12	AG125093	AG125093 Pan trogl
30	389.2	28.2	623	12	AQ037711	AQ037711 CTR-HSP-2
31	388.8	28.2	885	10	BG761979	BG761979 602718930
32	387.8	28.1	704	12	AG020199	AG020199 CTR-HSP-2
33	386.6	28.0	771	12	AG001229	AG001229 Homo sapi
34	383.4	27.8	671	12	AG045131	AG045131 Pan trogl
35	381	27.6	742	12	AQ748577	AQ748577 HS-5539_A
36	380.6	27.6	660	12	AG057341	AG057341 Pan trogl
37	380.2	27.5	686	12	AG119684	AG119684 Pan trogl
38	379.8	27.5	940	10	BG679573	BG679573 602627616
39	377	27.3	625	12	AG070563	AG070563 Pan trogl
40	376.6	27.3	739	12	AG001404	AG001404 Homo sapi
41	373.6	27.1	760	12	AG001225	AG001225 Homo sapi
42	373.6	27.1	858	9	AU122114	AU122114 AU122114
43	373.2	27.0	543	12	B94403	B94403 CTR-HSP-217
44	373.2	27.0	675	12	AG058688	AG058688 Pan trogl
45	369.6	26.8	746	12	AG001408	AG001408 Homo sapi

ALIGNMENTS

RESULT 1	BE420422	2615 bp	mRNA	linear	EST 11-OCT-2001
LOCUS	BE420422	32-393	human bone marrow CDNA	library	Homo sapiens CDNA, mRNA
DEFINITION	BE420422	sequence.			
ACCESSION	BE420422				
VERSION	BE420422.1	GI:16041640			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	Lu, X., Cui, L., and Li, Y.				
TITLE	DDRT-PCR from B cell				
JOURNAL	Unpublished (2000)				
COMMENT	Contact: xingwu lu, liangxian cui, yonghai li Department of Biochemistry Institute of Basic Medical Science, Peking Union Medical College Dongdan Sanliao 5, Beijing, P.R.C, 100005 Tel: 86-010-65296951 Email: luxingwu@263.net				
FEATURES	Location/Qualifiers				
source	1. 2615 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="human bone marrow CDNA library" /issue_type="bone marrow"				
BASE COUNT	705 a 531 c 641 g 738 t				
ORIGIN					
Query Match	49.1% Score 677.8; DB 10; Length 2615;				
Best Local Similarity	80.5% Pred. NO. 1.9e-182;				
Matches	859; Conservative 0; Mismatches 167; Indels 41; Gaps 4;				
Oy	352 gaatttggcgaatgaagaagaatttgctaaagaagaatttcgaagca 411				
Db	493 GAACTTGAGCTTGAGAGAGATCAITTAGGATATCTGCACAGAAATTTCTAAGCAGCA 552				

	Db	897	TTGTACTAG	906		
RESULT	4					
LOCUS	BK471183			1036 bp	mRNA	EST 05-FEB-2002
DEFINITION	AGENCOURT_6478314 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5563049					
ACCESSION	BK471183					
VERSION	BK471183.1					
KEYWORDS	EST.					
SOURCE	Homo sapiens					
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
REFERENCE	1 (bases 1 to 1036)					
AUTHORS	NIH-MGC http://mgi.nci.nih.gov/.					
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)					
JOURNAL	Unpublished (1999)					
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabbs@mail.nih.gov Tissue Procurement: ATCC/DCTP/dmp cDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov plate: LLMI2293 row: g column: 18 High quality sequence stop: 661. Location/Qualifiers 1. 1036					
FEATURES	source					
	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					
	/clone="IMAGE:5563049"					
	/clone_lib="NIH_MGC_72"					
	/tissue_type="melanotic melanoma"					
	/lab_host="DH10B (phage-resistant)"					
	/note="Organ: skin; Vector: pCMV-Sport6; site_1: NotI; site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2 kb. Library constructed by Life Technologies."					
BASE COUNT	254 a 239 c 294 g 247 t					2 others
ORIGIN						
Query Match	35.0%; Score 483; DB 10;					Length 1036;
Best Local Similarity	77.7%; Pred. No. 7,4e-127;					
Matches 676; Conservative	0; Mismatches 151; Indels 43; Gaps 6;					
Qy	554 gcgaaattgcatagtatagtagact--gaatttaattcctaagaacatggggaaa	611				
b	1 GCACAAATTAATACATAAGTAGACAAGCAAGTACTAATTGTTAAATCCCAAGAACATAGGGGA	60				
Qy	612 atatcttccttgacaattgcagaggcttcacagcagttcatcaaatcaactcgcttgagcc	671				
Db	61 ATGTCCTCAAGTTCCTGTGTGACAGACCTTCACTGCGACAGCCTTGCCATCACAGGCCAAGGC	120				
Qy	672 cttagag--aaaatggttttggagaaagcccaggggtccctgtgctgtgtgacgcta	728				
Db	121 CCAGAGAGAAGAAAGTGTGCTGTGGGGCTGGGTTCAGGGGTCCCATGCTGTGTGAAGCCTA	180				
Qy	729 gagacttgtgtgccctgtgtcccaagttaatta-----	760				
Db	181 GGAGCTTGTGGTCCCTGTGCCAGCTGTGCCAGCTTGCGTAGAAGGGGCTAATTGTATAG	240				
Qy	761 -----gctgtgagcttcagagggtgcaagcccccaagcctgtgcag---cttccaagtggtg	812				
Db	241 CTCGTGGCTGTGGCTTATGAGAGGTGAAGCCCCAAGCCTTGGCAGCTTCTTCCATGTGAGTG	300				
Qy	813 ttgagcctgtgggtgcagaaagcaagaatgaggttgggaacctccaatcagatttc	872				

Db	Accession	Source	Organism	Reference	Authors	Title	Journal
Db	301	TTGAGCTGTGGGTCACAGAGTCAAGAACTGAGGTTTGGGAACCTTGCT-AGATTCC	359				
Qy	873	agaaagtataatgtaaaaccccttgatgagcccaagcagaagtttgcctgtataggggtgggttc	932				
Db	360	AGAGAGATGATGAAATGCTTGATGGACCCAGCAAAAGTTTGCTGCAAGAGTTGGGCTCT	419				
Qy	933	catgagaaacctctgcaaggttagtacaaaagggaatgttgggtggagccccacaca	992				
Db	420	CATGGAGAACCTCTGCTAGGACGAGTGGAAGGAAATGTGGGTTTGAGCCCCCCACACA	479				
Qy	993	gagtcaccaatgggtggtccatctagtagagctgtgtgagaagaagtccaatcctccaagc	1052				
Db	480	GTGGCCCTGTGGGGGCACTGCTAGTGGAGCTGTGAGAGAGGGGCCACCATCCTCCAGAC	539				
Qy	1053	tccaagaaggttagatccatctgacagcttgagagcagctgtgctggaaaatccaacacac	1112				
Db	540	CCGAGATGATGATATCATCAACAGCTTGCACTGTGCACTTGGAAAGCCGACAGACATC	599				
Qy	1113	agtgcaagcctgtgaagaagcagagatgtgagctgtctacccatacaaaacgtatgtgcag	1172				
Db	600	GACACCAAGCCCATGAAAGCAGCTGGGAGGAGGAGCATATACCTGCAGAAACACAGGGGAG	659				
Qy	1173	agctgacccaagaccgtggggaatctaccccttgatctgtcattgtcagctgagagtgagcag	1232				
Db	660	AGCTGCTTAGACCAATGAGGAAACCCACCTCTTGCAATCAGCAGACACTGATATGAGACTTG	719				
Qy	1233	gagtcacaaaagatactatttggagctttaaagatttgactgcccactggaatttgcagct	1292				
Db	720	GAGTCAAAGAGATCAATTTTGGAGCTTTGGAATTTGATGCGCCCGGATTTCAACACTT	779				
Qy	1293	ataatggggccgct-accccttgttttggccaatttttccatttggagctcgctattt	1351				
Db	780	GTGATGCCCTGATATACCCCTTGTGTTTGGCAATTTGCCCATTTTGAAGGCTACATTT	839				
Qy	1352	accgaatgcctgactacatctcatgtgtatg	1381				
Db	840	TACCAATACCTGTACCCCTATGGATCTAG	869				
RESULT	5						
LOCUS	BI488505	870 bp	mRNA	linear	EST	28-AUG-2001	
DEFINITION	603021014F1 NIH_MGC_114 Homo sapiens CDNA clone IMAGE:5191803 5',						
ACCESSION	BI488505						
VERSION	BI488505.1	GI:15327733					
KEYWORDS	EST.						
SOURCE	human.						
ORGANISM	Homo sapiens						
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.						
TITLE	NIH-MGC http://mgc.nci.nih.gov/.						
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)						
COMMENT	Unpublished (1999)						
	Contact: Robert Strausberg, Ph.D.						
	Email: cgrabs-remail.nih.gov						
	Tissue Procurement: Life Technologies, Inc.						
	CDNA Library Preparation: Life Technologies, Inc.						
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)						
	DNA Sequencing by: Incyte Genomics, Inc.						
	Clone distribution: MGC clone distribution information can be						
	found through the I.M.A.G.E. Consortium/LLNL at:						
	http://image.llnl.gov						
	Plate: LLAM11479 row: k column: 04						
	High quality sequence stop: 849.						
FEATURES							
source							
	1..870						
	/organism="Homo sapiens"						
	/db_xref="taxon:9606"						
	/clone="IMAGE:5191803"						
	/clone_lib="NIH_MGC_114"						
	/lab_host="DH10B"						

/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27 yo. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.5 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 019. Note: this is a NIH_MGC Library."

BASE COUNT 222 a 211 c 256 g 181 t

ORIGIN

Query Match 34.7%; Score 479.2; DB 10; Length 870;
Best Local Similarity 76.1%; Pred. No. 8.5e-126;
Matches 645; Conservative 0; Mismatches 168; Indels 35; Gaps 3;

QY 493 gcaacctgacatgtagataaaagaaacccattctgagggggaattcaagctgac 552
DB 1 GCAGCCTGATGATGAAGTGAAGAAAAAACCATTTTGTGAGATTAATTCAGCCACG 60
QY 553 tgcagaanaattgcataatgtaataagagagctgtaattcaatcacaagaatgggaaaa 612
DB 61 TGCAGAAATTTGCATATGTAAGAGAGCTATGTTATCTCCAGACCATGGGAAAA 120
QY 613 tatctccgagacatgctcagaagcttcacagcagctccatcaatacactgcttgagac- 671
DB 121 TGTCTCCGGGCGCATGTCAAGACCTTCAAGGACCCCTCCATCACAGCGTGAGGCCA 180
QY 672 ctaggagaanaatggtttgt 731
DB 181 GGAGGAAAAAGAGGTTTGT 240
QY 732 actgtgtgacctgt 760
DB 241 ACTGTGTGCTGT 300
QY 761 --gcgt 818
DB 301 GGGGCTGT 360
QY 819 ctgt 878
DB 361 CTACAGGT 420
QY 879 tatatggaacccctgt 938
DB 421 TGT 480
QY 939 gaacctctgcaaggtagtaacaaaggaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 998
DB 481 GAACCTCTGT 540
QY 999 ccaagt 1058
DB 541 ATACGT 600
QY 1059 aggt 1118
DB 601 ATGT 660
QY 1119 agcctgt 1178
DB 661 AGCCATGT 720
QY 1179 ccaagacgt 1238
DB 721 CCAAGACATGT 780
QY 1239 aaagagacatttggagacttaa--gattgacgtgcccaagaggtttgagactatgtg 1297
DB 781 AAGGAGATCATTTTGAGACTTTAACAATGTGACCTGCTGCTGCTGCTGCTGCTGCTGCT 840

QY 1298 gggcccgct 1305
DB 841 GGGGCCCT 848

RESULT 6
LOCUS BM457166
DEFINITION BM457166 898 bp mRNA EST 05-FEB-2002
ACCESSION AGENCOURT_6411690 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5583427
VERSION BM457166
KEYWORDS 5', mRNA sequence.
SOURCE BM457166.1 GI:18506206
ORGANISM human.
REFERENCE NIH-MGC http://mgi.nci.nih.gov/
AUTHORS 1 (bases 1 to 898)
TITLE Mammalia; Euthera; Primates; Catarrhini; Homindae; Homo.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgrabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL12346 row: h column: 20
High quality sequence stop: 9
High quality sequence stop: 713.
Location/Qualifiers
1. 898
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5583427"
/clone_1ib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally: oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

BASE COUNT 225 a 228 c 254 g 191 t

ORIGIN

Query Match 34.6%; Score 477.8; DB 10; Length 898;
Best Local Similarity 77.3%; Pred. No. 2.2e-125;
Matches 669; Conservative 0; Mismatches 137; Indels 59; Gaps 5;

QY 563 tgcataatgtaagagctgtaattcctcaagaacatggaataatctctg 622
DB 35 TGCATTAAGCAGCAGAGAGGCTAATGTTATCCCAAGACCATGAGGAAATGCTCCAG 94
QY 623 acatgtcaagatcttcacagcagtcacatcaatcaactgtgctgagagcctgagag--a 679
DB 95 TCATGTCAAGACCTTTCATGAGCAGCCCTCCATCCACAGCCCGAGAGGAGGAA 154
QY 680 aatgt 739
DB 155 AAGT 780
QY 740 cctgt 766
DB 215 CCTGT 780
QY 767 gcttcaagaggtgtcaagcccaagcctgtgagccttcaagtgtgtgtgtgtgtgtgtgtgt 826
DB 275 GCTTCAGAGGTGAGAGGCCCAAGCATTTGGAGCTTGTGATGTTGAGCTGTGTGTGTGT 840

RESULT	11
LOCUS	BE728616
DEFINITION	BE728616 721 bp mRNA linear EST 15-SEP-2000
ACCESSION	601563827F1 NIH_MGC_20 Homo sapiens CDNA clone IMAGE:3833454 5'
VERSION	mRNA sequence.
KEYWORDS	BE728616 BE728616.1 GI:10142608
SOURCE	EST.
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE	Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
JOURNAL	1 (bases 1 to 721)
COMMENT	NIH-MGC http://mhc.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: ATCC/DCTD/DTF CDNA Library Preparation: Ling Hong/Rubin Laboratory DNA sequencing by: Incyte Genomics, Inc. clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: image.lnl.gov plate: LNCM512 row: e column: 07 High quality sequence stop: 714. Location/Qualifiers 1..721 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:3833454" /clone_1lb="NIH_MGC_20" /tissue_type="melanotic melanoma" /lab_host="DH10B (phage-resistant)" /note="Organ: skin; Vector: pOTB7; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGGCAGG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
FEATURES	
source	
BASE COUNT	185 a 180 c 205 g 151 t
ORIGIN	
Query Match	29.8%; Score 411; DB 10; Length 721;
Best Local Similarity	77.8%; Pred. No. 2.6e-106;
Matches 564; Conservative	0; Mismatches 120; Indels 41; Gaps 4;
Oy	aggagaaagtgttttgttggaacagcccaaggtcctgtgctgtgtgcagcctaagagac 733
Dd	4 AGGAATAATGTTCATGAGGGCGCCAGGCCAGGGTGCTCGATTGTGTACAGCGTAGAGAC 63
Oy	tltgtgccctgtgtccagtaattca----- 760
Dd	64 TTGGGCGCCCTGTGTCCAGCTTCCCACACTGTAGCTAAAGGGGCCAATGTACACTCAG 123
Oy	gtctgtgaccttgaagaagtgcaagcccccaagccttggagcttcgaagtggtgttagct 820
Dd	124 GCTGTGGCTTCAGAAGGTGAAGCCCCCAAGCCCTGTGACGCTTCATGTGTATTAGCCT 183
Oy	gttggtgtcgaagaagtcagaatttaggttttgggaacctccaatcagatttcaagaata 880
Dd	184 GGGGGTGCAAGAAATCAAGATTGTAGGTTGTGGGAACCTTAACGTAGATTTCAGAAAATG 243
Oy	tatggaaacccttggaatgcccagcagaagtttctgttaggggtgggttcccatgatgaga 940
Dd	244 TATGGAATATGCTTGGATGCCAGCGAAAAATTGCTGCTGAGAGAGTGGAGCCCATGAGAGA 303
Oy	accttgcgaaggtgtgtacaagaaggaaatgtttgttggtggagccccccacagaagttccc 1000
Dd	304 GCCTTGTGCTAGGGCGAGTGTGGAAGGAAATGTGGCGTTTGAACCCCATGAAAGTTCCT 363

QY	1001	agtggtggtccatctcagtagagctgtgtggaagaagatccacatcctctcagactccagaag	1060
Db	364	AGTGGGGCACTCCCTAAGTGAGACCTGTGAGAAAGGGCACTCTCCTCCAGACCCAGAT	423
QY	1061	gttagatccactgacagctgtgcagcatgtgcctgtgaaaaatccacagacactaagtgcag	1120
Db	424	GCTAATCCACTGCAGAGCTTGACCGCTGACCTGGAAAAACCGAGACCTCAATGCCAG	483
QY	1121	cctgtgaaagcagcagggatgtgagtcgtgtacccatacaaaacccgtagtgtgcagactgac	1180
Db	484	CC-ATGAAAGACGACAGGAGGGAGGCTGTACCTGCAAAAGCCACAGGGGACAGGCACCC	542
QY	1181	aagaccgtgtggaatctacactctgtcattgtcatgtcatgacctgtgcagatgagatgcaaa	1240
Db	543	GAGACCAATGGGAAACCACTCTAGCATAGTGTAACTGGACATGAGACCTGTGATTCAAA	602
QY	1241	agaagtcattgtgagacttaagaattgacgtccaccatgattccggaactatag-9	1299
Db	603	GGAGATTAATTTTGGAGCTTTAA-----AGTTGACCCCTGGATCTCAGACTTGCATTGGAC	656
QY	1300	gcccttaccccttctgtgtgccaatttttccatacttttgcattggaactgcgfatattaccatg	1359
Db	657	ACTGTAAACCCCTTGTTGTGGCCATTTCCTCCATTGTGAACAGCTATATTACCCAAA	716
QY	1360	cctgt 1364	
Db	717	CCTGT 721	
RESULT	12		
LOCUS	BG681769	955 bp	EST_01-MAY-2001
DEFINITION	602627877/1 NCI_CGAP_Skn4 Homo sapiens CDNA clone IMAGE:4752858	5'	
ACCESSION	BG681769		
VERSION	BG681769.1	GI:13913153	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 955)		
TITLE	NH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs@mail.nih.gov Tissue Procurement: James Cleaver, M.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC Clone distribution/LNLN at: http://image.lnl.gov Plate: LNLN10611 row: 1 column: 19 High quality sequence stop: 793.		
FEATURES	location/qualifiers		
Source	1..955		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:4752858"		
	/clone_1id="NCI CGAP_Skn4"		
	/lib_host="squamous cell carcinoma"		
	/lib_type="DH10B (T1 phage-resistant)"		
	/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NciT; Site_2: SalI; Cloned unidirectionally. Primer: oligo dT. Average insert size 1.5kb. Library constructed by Life Technologies. Note: this is a NCI_CGAP Library."		
BASE COUNT	270 a 216 c 273 g 196 t		
ORIGIN			

Query Match 29.5%; Score 407.8; DB 10; Length 955;
Best Local Similarity 76.9%; Pred. No. 2,3e-105;
Matches 645; Conservative 0; Mismatches 142; Indels 52; Gaps 10;

Qy 397 aattcttaagcagcaaaacattcaagaagtgactgggtgctctttaagcattcgttt 456
Db 1 AATTTTCTTAAGCAGCAGAGCATTCACAGAGTGTGGGTCTGTATTAAGCATTCAGTTT 60

Qy 457 cataagggagagagacataagactcagaaatttgaccctgacacatctgataaaaa 516
Db 61 TAAAGAGAAACAGAGCATTAAGAGTTGGAGATTTTCACCTGACATGACATAGTAAGAAA 120

Qy 517 gaaaaaccatttcttgagggaattcaagcctgctgagaaatttgcatatgtaata 576
Db 121 GAAAA-----TCTCAGGTGACATTCAGAGCTGACAGAGTGTGATTAAGTAATGA 172

Qy 577 gggagcgaattttaaactcctaagaacataatgggaaataatctcttgacatgtcagaagtc 636
Db 173 GGGGCGAAATGTTAAATCCCAAGAAAATGGGAAAATGTCTCCAGGGCATGTGACAGATC 232

Qy 637 ttacacagagtcacatcaaatcactgctgagagcctagag--aaatggtttgtgg 694
Db 233 TTGATGAGCAGCCCTCCATCAGAGCCTGACAGCCTAGAGAGAAAAGGTTTCATGG 292

Qy 695 aagagcccaagggtccctgtgctgtgctgagcctagagactggtgctgctgtgccagt 754
Db 293 GCGAGGCCAGAGTCCCATGCTGTGAGAGCTAGAGACTTGGTGCCTGCATCTCAGCC 352

Qy 755 aattcagc-----ttggtcttcagaagggtga 781
Db 353 GCTCTTAGCCATGAGCAAAAAGGGGCCAACGTAGACCTAGCATGTGGCTTCAGAGGATGA 412

Qy 782 agcccaagccttggcagcttcccaagtggtgttgagcctgtgggtgcaagaagtcaaga 841
Db 413 AGGCCCAAGCCTTGGAGAGCTT-CACATGCTGTGAGCCTGCAAGTGACATTAAGTAAGA 471

Qy 842 attgaggtttgggaacctcccaatcagatttcaagaatataatggaaacccctgagtcgc 901
Db 472 ACTGGGGTTTGGGAMCTCCACCTAGATTTCAGAAAGATGATGAATGTCTGGATGCC 531

Qy 902 aggcagaagtttgcgtgtaggggtggtgctcctcagtagaacctctcagaaggtgaat-aca 960
Db 532 AGGCAAGAA-TTGTCTGTCATGGGAGGCGCCCTTATGGAAGACCTCTAC TAGGGCAGTGTGA 590

Qy 961 aaggggaatgtgtgtgtggagccccccacacagagttcccaagtggggctccatctagtag 1020
Db 591 AAGAGGAATGTGAGGTGGTGGTGGCCCCCAGACAGAGTTCCTACAGGAGCACTGCTAATGG 650

Qy 1021 agctgtgagaaagatcccaatcctccagactccagagctcagaaggtgagtagtccatctgacagc--t 1079
Db 651 AGCTGTGAGAAAGAGAGGCCAGCATCTCCAGATCCAGAAATGATCCACTACAGCTT 710

Qy 1080 tgcagatgtgccttgaataatccagagactagtgacagcctgtgaaagcagcagggga 1139
Db 711 TGCACACGTCGCTGGAAGAGCTGCAAGCATTCAGACCAAG-CTGTGAAGACGACA-6GA 768

Qy 1140 tggagctgtacccaacaaaccgtagctgagagctgacacaagcgtgagtagtctac 1198
Db 769 GGGAAACTGTACTGTAGAGCAGC---AGGGTAGAGCTGCGCCAGAGACATGAGGAAACAC 824

RESULT 13
BG281182 760 bp mRNA linear EST 21-FEB-2001
LOCUS 602402107P1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:454448 5',
DEFINITION mRNA sequence.
ACCESSION BG281182
VERSION BG281182.1 GI:13030107
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 760)
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC/DCTD/DTP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMNL at: image.llnl.gov
Plate: L10M127 row: n column: 01
High quality sequence stop: 760.
Location/Qualifiers
1..760
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:454448"
/clone_1b="NIH_MGC_20"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pORF7; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(6). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using 2AP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 211 a 181 c 219 g 149 t
ORIGIN

Query Match 29.4%; Score 405.4; DB 10; Length 760;
Best Local Similarity 77.5%; Pred. No. 1e-104;
Matches 571; Conservative 0; Mismatches 126; Indels 40; Gaps 5;

Qy 444 aggcattcagttcttcaagggagcagacataagattcaagaataattgacacctgaca 503
Db 2 AGGCATTCAGTGTATTAAGGA--ACACGATMAAGTTGACGAGCCGATG 59

Qy 504 atgtgataaaaaaagaaacccattcttgagggaattcaagctgctcagaaattt 563
Db 60 ATCCAGTGAAGAAAGAAAACCATTTTGTAGGAAAAAATTCAAGCCAGCTGACGAATTT 119

Qy 564 ggaattgtaagtagagagtgatattatcctcaagaacaatgggaaatatctcctgga 623
Db 120 GCGAAGTAGCAAGAGCTTAAATGTTAAATCCCAAGACATTAAGGAAATATCTGCAAGC 179

Qy 624 catgtcaaggtcttcaagcagctcacaatcaactggtcctgagagccttagag--aa 680
Db 180 CATGTCAAGAGACTTCANAGCAGCCCTCCCATCAGAGGCCAGAGGCCAGAGAGAAA 239

Qy 681 aatggttttgaggacagggccagaggtccctgtgctgtgtgcaagccttagagactgtgtgc 740
Db 240 AGTGGTTTCATGGGCCAAGGCCAAGGTCCTTATGCTGTGTGAGCTAGCGACTGTTGGTC 299

Qy 741 cctgtgtccagttaatca-----gctgtgg 767
Db 300 GTTGTATTCACCCACTCCACCATGAGCTGAAGGGCCAAATGTAAAGACTCGGGTGTGG 359

Qy 768 ctccagaggtgcaagcccaagccttgagccttccaaigtgttgagcctgtgagtg 827
Db 360 CTTTCAGAGGGTGAAGGCCCAAGCTGTGGCAAGCTTCATATGTGTGTTCAGCTGCAAGTGG 419

Qy 828 caaagaagtcagaattgaggtttgggaacctccaatcagattcagaagatatatgaa 887
Db 420 CACAGAAATCAGAAATTAAGGTTGGGAACCTCCACTGATTTGAGAGGA-GTATGAAA 478

Qy 888 acccctgagtcgccagaggaaggtttgtctgttaggggtgggtccccaatgaggaactctg 947
Db 479 ATGCTTGATGCCAGCGAAGATTTGTCTGAGGCGAGAGGCCCTCAAGGATTAACCTGTG 538

QY	948	caagaggtgtgtacaaagaagaaatgctgtgtgtgtgagagccccacacagaagtccccagttgagg	1007
Db	539	CTAAGGGCAGATCGAGAAAGGAAATGTGGGATTAGACCCCTGACACAGATCCCCCACTGGGG	598
QY	1008	ctccatctcagtagaagctgtgagaagaagtccacatctccagaactccagaaggttagat	1067
Db	599	CTCTCCCTTAGTGGAGCCTGTGAGACAGC-GCACCACTTCTCCGATCCCGATGGTGTAGT	657
QY	1068	ccactgtaacagctcttgagcatgtgcctgaaaaatccacagacacactcagtcagccgctgtga	1127
Db	658	CCACGCGAGCGTTTGCATCATGCACCTGAGAAAGACAGACACTCAATGCCACCTGTGA	717
QY	1128	aagcagcagaggaatgag	1144
Db	718	AAGCAGCCAGGAGGAG	734

RESULT	14	736 bp	mRNA	linear	EST 25-SEP-2001
LOCUS	BT754555/c				
DEFINITION	603023384F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5194028 5',				
ACCESSION	BT754555				
VERSION	BT754555				
KEYWORDS	BT754555.1 GI:15746133				
SOURCE	EST.				
ORGANISM	human.				
	Homo sapiens				

REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 736)
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: LHAM1485 row: 9 column: 21
High quality sequence stop: 733.

FEATURES	Location/Qualifiers
source	1. .736

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5194028"
/clone_id="NIH_MGC_114"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-Sport6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27 yo. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.5 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 019. Note: this is a NIH_MGC Library."

```

Query Match	29.3%	Score 404.6;	DB 10;	Length 736;
Best Local Similarity	76.6%	Pred No. 1.8e-104;		
Matches 563; Conservative	0;	Mismatches 134;	Indels 38;	Gaps 4;
OY	655	atcactgacctgagagcctagag--aaatgylttgtgagacagcccaaggtccct	711	
733	ATCACAGACCCAGAGAGCTAGACGCAAAAAATGCTTTTGTGTCTGACGCCAAGTCCCC	674		

OY	712	gtctgtgtgtcagcctaaagaacttvgtgccctgtgtcccaattaatctagcttggc---	768
Db	673	ATGTTGTGTGACGCTAG-GACTTGCTCCTTGTTCCACCCACTCAGGCATTGGCTAA	615
OY	769	-----ttcaagggtgcaagccccaagccttgcga	798
Db	614	AAGGGGCCAAGGTACAGCTCAGCCCATGTTTCAAGAGTGCMAACCCAA-CCTTGGCA	556
OY	799	gcttccaagtggtgttgaagcctgtgtgtcagaagaagtcagaatctgaagtttggaaac	858
Db	555	GCTACACATGTGTGTTGAGCCTGGCGGTGTACAGAACTCAAGAAATTGAGTTTGGGAACC	496
OY	859	ttccaatcagatttcaagaatatataggaaacccctggatgtccagggcagaagtttgcgt	918
Db	495	TCCATCTTAGATTTCAGAAAGATGTATGAAAAAGCCTGGAGTCCCGCAAAAGTTTGCTGC	436
OY	919	agggtgtgggtctcatttgaagaaacctctgcagaagtgatlaacaaaaggaaatgttgggtg	978
Db	435	AAGGGAGGGGGCCCTTGTGGAGAACCTCTGTATGGGCAATGCAGAAAGGAATGTGGGGTCC	376
OY	979	ggaggccccaacagagagccccagtgggggcccaatctgtatgagctgtggaagaagaagtc	1038
Db	375	AGAGCCCCAACAACAGTATCCCTACTGGGGTACTGGCTTAGTAGAGCTGTAAAGAAAGGCC	316
OY	1039	accatctctccagactccagaaggttagatctcaacttgcacttgcagcatgttgccttgaaa	1098
Db	315	ACCATCCCTCCAAACATCAAGAAATGTAGATATACATGACAGCTTGTAACAAGGTGCTGGAAA	256
OY	1099	atcccaagaacactcagtcagccagcctgtgaagaagcagagtgtagtcttgaacctaa	1158
Db	255	AGCCGCAACACTCATATCCACGCCCATTAAGAAAMCAGCAGGGAGAGCTATGCCCTGCNA	196
OY	1159	aaccgttagtgagagcttgcacaaagaccgttggaaatctaccctctgcatttgcagact	1218
Db	195	AGCCAGCAGGGGTGGAGTGCCTACAGACTATGGGAACAACCTCTTGATCACTGAGACTT	136
OY	1219	ggacgttggagaatcagtagtctcaaaagagatcatttggagctttaaagatttgcgtcccaac	1278
Db	135	GGATTTGGAATATGAGTCAAAGAAATCACTTTTGGACCTTCAAAATTTTGATGCTGCCCGC	76
OY	1279	tgaatttcgaaactaatagggtgcccgtaccccttgttcttgcacaaatttcttcatttgg	1338
Db	75	TGGATTTTGGACTTGGCAATGGGGTCTGTAAACCCCTTTGTTTGGGCATTTTCTCCATTGGG	16
OY	1339	aactgcgatttacc 1353	
Db	15	AATGGCCGTATTTCAC 1	

REFERENCE	ORGANISM	KEYWORDS	VERSION	ACCESSION	LOCUS	RESULT 15
AG062447						
AG062447						
AG062447						
AG062447.1						
GI:16614249						
691 bp						
DNA						
linear						
GSS 03-NOV-2001						
genomic survey sequence.						
Pan troglodytes male lymphoblast DNA, clone_11b:PTB Chimpanzee Male						
BAC library clone:PTB-050M20.F.						
Pan troglodytes						
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.						
1 (sites)						
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,						
Totoki, Y., Watanabe, H., and Sakaki, Y.						
BAC end sequences of library PTB						
Unpublished						
2 (bases 1 to 691)						
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,						
Totoki, Y., Watanabe, H., and Sakaki, Y.						
Direct Submission						
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical						
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);						
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan						

COMMENT

(E-mail:chimpbes@sc.riken.go.jp, URL:http://hnp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the Rad process and may have higher chance of
clone tracking errors.

PRIMERS

Sequencing: -21M13

LIBRARY

Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.

FEATURES

source
1. .691
Location/Qualifiers
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-050M20.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC library"
BASE COUNT 160 a 167 c 201 g 162 t 1 others
ORIGIN

Query Match 29.3%; Score 404.2; DB 12; Length 691;

Best Local Similarity 82.1%; Pred. No. 2.2e-104;

Matches 503; Conservative 0; Mismatches 103; Indels 7; Gaps 3;

```
QY 746 gtccagtaattcagctgtgcttcagaggtgcagagcccaagccttgcagcttcca 805
    ||||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 83 gtccggaggtcagagctgtgcttcagaggtgcagagcccaagccttgcagcttcca 142
QY 806 agtgggttgagagcctgtgtgcaagaagtgcagagtttgagagcttgcacccaac 865
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 143 tctgtgttgagacctgtgagtgcaag-----aaagatcgaggtttggagaccttcacatt 197
QY 866 agattcagaagatataatgaacccctggatgcagagcagagatttgcgtgaagggtg 925
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 198 agatttcagagagatttatgaaatgactgagtgcccatgcgcaaaatttctgtaggtca 257
QY 926 gggctcctatgagagacctctgcagaggtgtacaagaaggaattgttggtggagccc 985
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 258 gggcctctatgagagacctctgtgaggtgaggtgagaggaatgtgaggtcagagggcc 317
QY 986 ccacacagagtcgccagtggtggtccatctagtagagctgtgagaagaagtcacacatcc 1045
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 318 ccacacagagtcgcttactgaggggacacacccagtgagctgtgagaagagggccactgtcc 377
QY 1046 tccagactccagaaggttagatccactgcagacttgcagcatgtgcctgaaaaatccaca 1105
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 378 tccagactccagaaggttagatccactgcagacttgcagcatgtgcctgaaaaatccaca 437
QY 1106 gaacactcagtcagagcctgtgaaagcagcagagatgagtcgttacccctacaagaacgta 1165
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 438 gaacactcagtcagagcctgtgaaagcagcagagatgagtcgttacccctacaagaacgta 497
QY 1166 gtggcagaagctg-accaagacggtgggaatctaacctctgtcatgtcagacctgagcgt 1224
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 498 gggcgacagactgttccaaagaccatgggaaccacaccttgcctcagtgtagctgtgtgt 557
QY 1225 gaagacatggagtcaaaagaagatcatatttgagccttaagaattgactgcccacitgatt 1284
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 558 gaagatgagagtcagagagagatcatatttgagcctttaaagattgactgtgtggcttgatt 617
QY 1285 tggagcttataggggcccgta-cccccttgattttggcgaatttttccatttggaaactg 1343
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 618 tgggaccttgacatggggccctgtagctcttctgttttgcccaatttctccattttgtaattgg 677
QY 1344 ccgtatttaacca 1356
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 678 ctgtattttaccga 690
```

Search completed: June 30, 2002, 14:02:47
Job time: 8074 sec

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Mp: MORNREP; Information on the MORNREP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormrep This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/MRP/Chr22>
RP1-302D9 is from the library RPCT-1 constructed at the Roswell Park Cancer Institute by the group of Pletier de Jong. For further details see <http://bacpac.med.buffalo.edu/>

VECTOR: PCYPAC2
This sequence is the entire insert of clone RP1-302D9 The true left end of clone CTR-41562 is at 69682 in this sequence. The true right end of clone CTR-41562 is at 55167 in this sequence.

FEATURES

source

1. 145880

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="22"

/clone="RP1-302D9"

/clone_lib="RPCT-1"

188..245

repeat_region

/note="MER3 repeat: matches 144..209 of consensus"

246..571

repeat_region

/note="Alusx repeat: matches 1..312 of consensus"

572..759

repeat_region

/note="MER3 repeat: matches 1..144 of consensus"

783..933

repeat_region

/note="MERSA repeat: matches 26..187 of consensus"

1033..1336

repeat_region

/note="Alusp repeat: matches 1..299 of consensus"

1450..1583

repeat_region

/note="MIR repeat: matches 24..160 of consensus"

1687..1752

repeat_region

/note="L2 repeat: matches 2593..2661 of consensus"

2350..2660

repeat_region

/note="Alusc repeat: matches 3..309 of consensus"

2684..2981

repeat_region

/note="Alusg repeat: matches 2..300 of consensus"

3323..3343

repeat_region

/note="MTRIE repeat: matches 116..136 of consensus"

3344..3652

repeat_region

/note="Aluy repeat: matches 1..309 of consensus"

3653..3928

repeat_region

/note="MTRIE repeat: matches 136..359 of consensus"

3929..4278

repeat_region

/note="THEIB repeat: matches 3..364 of consensus"

4279..4485

repeat_region

/note="MTRIE repeat: matches 359..568 of consensus"

5073..5176

repeat_region

/note="52 copies 2 mer ct 78 conserved"

5181..5491

repeat_region

/note="Alubj repeat: matches 1..311 of consensus"

6369..6485

repeat_region

/note="L2 repeat: matches 2579..2705 of consensus"

6647..6685

repeat_region

/note="MADE1 repeat: matches 1..23 of consensus"

6686..6987

repeat_region

/note="Alusx repeat: matches 1..302 of consensus"

6988..7036

repeat_region

/note="MADE1 repeat: matches 23..77 of consensus"

7482..7754

repeat_region

/note="Alubj repeat: matches 9..290 of consensus"

7482..7754

repeat_region

/note="Alubj repeat: matches 9..290 of consensus"

7482..7754

repeat_region

repeat_region

775..8060

/note="Alujo repeat: matches 1..295 of consensus"

8414..8551

repeat_region

/note="L2 repeat: matches 2553..2706 of consensus"

8914..9030

repeat_region

/note="MIR repeat: matches 147..262 of consensus"

9110..9280

repeat_region

/note="MIR repeat: matches 91..262 of consensus"

9283..9412

repeat_region

repeat_region

/note="MIR repeat: matches 15..144 of consensus"

9521..9679

repeat_region

/note="FAM repeat: matches 3..161 of consensus"

9820..10225

repeat_region

repeat_region

/note="MSTRB repeat: matches 2..425 of consensus"

10179..10678

misc_feature

/note="match: GSS: Em:B56592"

complement(10204..10728)

misc_feature

misc_feature

/note="match: GSS: Em:A0701486"

complement(10249..10706)

misc_feature

/note="match: GSS: Em:A0225495"

10312..10383

repeat_region

repeat_region

/note="MIR repeat: matches 79..150 of consensus"

10718..11310

misc_feature

/note="match: GSS: Em:BI4024"

10784..11201

repeat_region

repeat_region

/note="match: GSS: Em:B43656"

11838..11946

repeat_region

/note="MIR repeat: matches 20..137 of consensus"

12174..12445

repeat_region

repeat_region

/note="L2 repeat: matches 1988..2275 of consensus"

12444..12642

repeat_region

/note="MIR repeat: matches 63..241 of consensus"

13017..13369

repeat_region

repeat_region

/note="match: STS: Em:G49301"

13331..13397

repeat_region

/note="MIR repeat: matches 174..244 of consensus"

13398..13698

repeat_region

repeat_region

/note="Alusp repeat: matches 1..302 of consensus"

13699..13810

repeat_region

/note="MIR repeat: matches 76..174 of consensus"

13806..13919

repeat_region

repeat_region

/note="MIR repeat: matches 77..189 of consensus"

13945..14060

repeat_region

/note="MIR repeat: matches 24..142 of consensus"

14061..14367

repeat_region

repeat_region

/note="Aluy repeat: matches 1..301 of consensus"

14368..14452

repeat_region

/note="MIR repeat: matches 141..225 of consensus"

14589..14679

repeat_region

repeat_region

/note="MIR repeat: matches 173..262 of consensus"

14597..15201

misc_feature

/note="match: GSS: Em:A053482"

14616..15060

misc_feature

misc_feature

/note="match: GSS: Em:AQ370601"

14868..15040

repeat_region

/note="MIR repeat: matches 49..233 of consensus"

15071..15188

repeat_region

repeat_region

/note="L2 repeat: matches 2112..2239 of consensus"

15304..15399

repeat_region

/note="MTRIB repeat: matches 1..99 of consensus"

15490..15662

repeat_region

repeat_region

/note="Alusg1 repeat: matches 2..114 of consensus"

15669..15727

repeat_region

/note="MTRIB repeat: matches 119..178 of consensus"

15728..16027

repeat_region

repeat_region

/note="Alusx repeat: matches 1..299 of consensus"

16028..16245

repeat_region

/note="MTRIB repeat: matches 178..390 of consensus"

16546..16854

repeat_region

repeat_region

/note="Aluy repeat: matches 1..300 of consensus"

18296..18323

repeat_region

/note="MSTR repeat: matches 2..29 of consensus"

18324..18392

repeat_region

repeat_region

/note="MSTR repeat: matches 2..29 of consensus"

18324..18392

repeat_region

/note="MSTR repeat: matches 2..29 of consensus"

18324..18392

repeat_region

repeat_region

/note="MSTR repeat: matches 2..29 of consensus"

18324..18392

repeat_region

/note="MSTR repeat: matches 2..29 of consensus"

18324..18392

repeat_region

repeat_region

/note="MSTR repeat: matches 2..29 of consensus"

18324..18392

repeat_region

/note="MSTR repeat: matches 2..29 of consensus"

18324..18392

repeat_region

repeat_region

/note="MSTR repeat: matches 2..29 of consensus"

18324..18392

repeat_region

/note="MSTR repeat: matches 2..29 of consensus"

18324..18392

repeat_region

repeat_region

/note="MSTR repeat: matches 2..29 of consensus"

18324..18392

repeat_region

/note="MSTR repeat: matches 2..29 of consensus"

18324..18392

repeat_region

repeat_region

/note="MSTR repeat: matches 2..29 of consensus"

18324..18392

repeat_region

/note="MSTR repeat: matches 2..29 of consensus"

18324..18392

repeat_region

repeat_region

/note="MSTR repeat: matches 2..29 of consensus"

18324..18392

repeat_region

/note="MSTR repeat: matches 2..29 of consensus"

18324..18392

repeat_region

repeat_region

/note="MSTR repeat: matches 2..29 of consensus"

18324..18392

repeat_region

/note="MSTR repeat: matches 2..29 of consensus"

18324..18392

repeat_region

repeat_region

/note="MSTR repeat: matches 2..29 of consensus"

18324..18392

repeat_region

/note="MSTR repeat: matches 2..29 of consensus"

18324..18392

repeat_region

repeat_region

/note="MSTR repeat: matches 2..29 of consensus"

18324..18392

repeat_region

/note="MSTR repeat: matches 2..29 of consensus"

18324..18392

repeat_region

repeat_region

/note="MSTR repeat: matches 2..29 of consensus"

18324..18392

repeat_region

/note="MSTR repeat: matches 2..29 of consensus"

18324..18392

repeat_region

repeat_region

/note="MSTR repeat: matches 2..29 of consensus"

```

repeat_region /note="MER66-internal repeat: matches 4919. .4993 of
consensus"
18393. .18712
/note="Aluub repeat: matches 1. .311 of consensus"
18713. .19133
/note="MER66-internal repeat: matches 4548. .4919 of
consensus"
complement(18872. .19230)
/note="match: GSS: Em: AQQ05063"
19251. .19719
/misc_feature /note="match: GSS: Em: B14179"
19537. .20230
repeat_region /note="HERVH21 repeat: matches 4657. .5784 of consensus"
20317. .20382
repeat_region /note="33 copies 2 mer ta 68 conserved"
20513. .20666
repeat_region /note="77 copies 2 mer tt 70 conserved"
20682. .21008
repeat_region /note="AluSg1 repeat: matches 1. .306 of consensus"
21239. .21553
repeat_region /note="HUES-P3 repeat: matches 4410. .4713 of consensus"
21882. .22254
repeat_region /note="THE1B repeat: matches 1. .364 of consensus"
22302. .22537
repeat_region /note="MER66-internal repeat: matches 2186. .2417 of
consensus"
22538. .22850
repeat_region /note="AluSp repeat: matches 1. .313 of consensus"
22851. .23801
repeat_region /note="MER66-internal repeat: matches 1210. .2186 of
consensus"
23905. .23989
repeat_region /note="MER66-internal repeat: matches 3017. .3102 of

```

Query Match	93.0%;	Score 1284;	DB 9;	Length 145860;
Best Local Similarity	99.6%;	Pred. No. 0;		
Matches 1287;	Conservative	0;	Mismatches 5;	Indels 0;
				Gaps 0;

OY	90	tgaggccctctggtgacacaggtttacacaaatatacaggagaaataagtgaatgacaa	149
Db	37254	TGGAAATGTCCTGGTCCCCAGGTTTACCACAATATACAGGAGAAATATAGTGAATGACAA	37313
OY	150	aatgcctctgtctcgtatataagaaagtcagccttactctgtagcgtggaatacttc	209
Db	37314	AATGCCCTGTCCTATATATAGAAAGGTACAGCTTACTGTGAAGCTACGTGAAACTTC	37373
OY	210	cccttcctttcaagcccaatcattcaagaaggctctgtacaaatcccaagaggaattaa	269
Db	37374	CTCTTCCTTTCAAGCCCATCATCTTCAAGGGGTCTGTACAAATGCCAGAGGATTTAA	37433
OY	270	aggagcgatggaggtcttctgcttgcagagtgccctgggaattactactccagctttag	329
Db	37434	AGGAGGCCATGGAGACTCTTGGCTTCAGAGGTCCCTGGGAATATCACTCCACTTTGANG	37493
OY	330	ttgagctcattcatcgcaaggtgaatttggctataagagaagaaatttggctata	389
Db	37494	TTGACTCTCATATTGCAAGGAGATATTTGGCTATATGAGCAAGCAAAATTTGGCTAATA	37553
OY	390	aggaaagaatttctaagcagcaaaagcattcaagagtgactcgggtgtcgttaaaagcat	449
Db	37554	AGGAAGAATTTCTTAAGCAGCAAAACCATTTCAAGAGGTGACTTGGGTGCTTTAAAGCAT	37613
OY	450	tcagttccaataaggagagcagagcaataagagttcagaanaatttcgacccttgacaa	509
Db	37614	TCAGTTTATTAAGGAGAGCAGAGCAATTAAGAGTTTCAGAAAATTTGCACCTTCACATGTGA	37673
OY	510	taaaaaagaaaaaacccatttcttgaggggaaattcaagctggtcgcagaaatttcatat	569
Db	37674	TAAAAAAGAAAAACCCATTTTCTGAGGGGAAATTTAAAGCTGGCTCAGAAATTTGCATAT	37733
OY	570	gtaatggagagctgaagttaatccctcaagaacaatlgggaaataatctctcgtagac	629
Db	37734	GTAATGAGAGAGCTGAATTTATCTCTCAAGCAATAGGGGAAATTTCTCTGACATGTC	37793

[illegible]


```
FEATURES
source
Location/Qualifiers
1..40714
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="X"
/clone="L10XNC01-212C1"
/clone_id="L10XNC01"
10..233
/note="L1 element fragment"
repeat_region
388..758
/note="L1 element fragment"
repeat_region
932..1108
/note="L1 element fragment"
repeat_region
2571..2617
/note="MER43 element fragment"
repeat_region
2613..5203
/note="L1 element fragment"
repeat_region
5204..5229
/note="L1 element fragment"
repeat_region
5230..5427
/note="MER43 element fragment"
repeat_region
8954..9088
/note="L1 element fragment"
repeat_region
12171..13985
/note="L1 element fragment"
repeat_region
14179..14612
/note="L1 element fragment"
repeat_region
14662..14780
/note="L1 element fragment"
repeat_region
14824..15044
/note="L1 element fragment"
repeat_region
15234..15369
/note="L1 element fragment"
repeat_region
15421..15849
/note="L1 element fragment"
repeat_region
16026..16155
/note="MSTR element fragment"
repeat_region
16073..16149
/note="THE1B element fragment"
repeat_region
16204..16495
/note="Alu repeat: matches 1..308 of consensus"
repeat_region
16532..16609
/note="THE1B element fragment"
repeat_region
16624..16722
/note="MSTR element fragment"
repeat_region
16626..16737
/note="MSTR element fragment"
repeat_region
16651..16712
/note="THE1B element fragment"
repeat_region
17060..17653
/note="L1 element fragment"
repeat_region
19720..19797
/note="MER21B element fragment"
repeat_region
19877..19915
/note="MER21B element fragment"
repeat_region
20007..20256
/note="MER21B element fragment"
repeat_region
20518..20734
/note="MER30 element fragment"
repeat_region
22661..22730
/note="MLT1A element fragment"
repeat_region
22740..22814
/note="MLT1A element fragment"
repeat_region
22917..22974
/note="MLT1A element fragment"
repeat_region
23342..23501
/note="THE1B element fragment"
repeat_region
23550..23697
/note="THE1B element fragment"
repeat_region
23725..24729
/note="THR element fragment"
repeat_region
24395..24593
```

```
repeat_region
2463..25270
/note="THR element fragment"
repeat_region
25277..25463
/note="THE1B element fragment"
repeat_region
25528..25627
/note="THE1B element fragment"
repeat_region
26081..26196
/note="MLT1A element fragment"
repeat_region
26265..26661
/note="L1 element fragment"
repeat_region
26797..26898
/note="MSTR element fragment"
repeat_region
27949..28241
/note="Alu repeat: matches 308..1 of consensus"
repeat_region
28319..28608
/partial
/note="Alu repeat: matches 308..1 of consensus"
repeat_region
30032..30159
/note="MER5 element fragment"
repeat_region
31794..31980
/partial
/note="Alu repeat: matches 308..109 of consensus"
repeat_region
31981..32073
/partial
/note="Alu repeat: matches 93..1 of consensus"
repeat_region
32232..32735
/note="MER9 element fragment"
repeat_region
35168..35654
/note="MER4B element fragment"
repeat_region
40640..40693
/note="27 copies of 2 mer 94 & conserved"
BASE COUNT 11544 a 7879 c 8141 g 13150 t
ORIGIN
Query Match 50.2%; Score 692.6; DB 9; Length 40714;
Best Local Similarity 80.6%; Pred. No. 1.8e-191;
Matches 873; Conservative 0; Mismatches 174; Indels 36; Gaps 4;
QY 335 ctgcatactatgcaagtgatatttggcctaagagagaagaatttgcataaaga 394
|||||
DB 24003 CTGCCCTAGAGATCTGTGGAAATTTGACTTGAGAGAGATTTAGGATATGAGTGGA 24062
QY 395 gaatttctaagcaagaagcatccaagagtgacttggtgctgttcaaggaattcagt 454
|||||
DB 24063 GAAATTTCTAAGCAGCAAGCATTTCAAGAGTGACTTGCTCTGTTAAAGCAGTCACT 24122
QY 455 ttcataaggagagcagagatgaagttcagaanaattgcacctgacaaatgtgataaa 514
|||||
DB 24123 TTTCGAAGGAGAGAGAGAACATTAAGTTTCAGCAAAATTTTCACCTGACAAATGTGATAGAA 24182
QY 515 aagaanaaacca-ttttcgagggagaattcaagctggctgaggaatttgcataatgtta 573
|||||
DB 24183 AAGAAACCCCATTTTTCGATGATTAATGAAGCCAGTGAGAAATTTGGATGAATGA 24242
QY 574 tgaggagctgaatgttaattccctcaagacaaatggggaanaatctccctgacatgcaag 633
|||||
DB 24243 CAGAGAGCCAAATGTTAATCACTAAGACCATGGGAAAAATATCTCATGTGCTGACAG 24302
QY 634 gtctcacagcagtcacatcaaatcaactg-gcctggagggcctagagaaatgtttgtg 692
|||||
DB 24303 GCTTTCAGGGAGGCCCTCATACATCTGCCAGAGGCCCTGAGGAAATGTTTATG 24362
QY 693 ggaacagcccaagggtccctgtgctgtgtgacagcctagagacttggtccctgtgccag 752
|||||
DB 24363 GCTCGGGCCAGAGGTCCTCATCTATGTGCACTTAAGGAGGATGTGGCCCTGTCAG 24422
QY 753 ttaattca-----gcttggtcttaaggaggtg 779
|||||
DB 24423 CCACCTCAGCCATTAACCTGAAGGGCCAAAGATTAAGCTTGCGCTGTGCTTCAAGGATG 24482
```


*	19528	18627:	gap of	100	bp
*	18628	20561:	contig of 1934	bp	in length
*	20562	20661:	gap of	100	bp
*	20662	22484:	contig of 1822	bp	in length
*	22485	22584:	gap of	100	bp
*	22585	22587:	contig of 2703	bp	in length
*	22588	22587:	gap of	100	bp
*	22588	27603:	contig of 2216	bp	in length
*	27604	27703:	gap of	100	bp
*	27704	30020:	contig of 2317	bp	in length
*	30021	30120:	gap of	100	bp
*	30121	33472:	contig of 3352	bp	in length
*	33473	33572:	gap of	100	bp
*	33473	38084:	contig of 4512	bp	in length
*	38085	38184:	gap of	100	bp
*	38185	41069:	contig of 2885	bp	in length
*	41070	41169:	gap of	100	bp
*	41170	45266:	contig of 4097	bp	in length
*	45267	45366:	gap of	100	bp
*	45367	51092:	contig of 5726	bp	in length
*	51093	51192:	gap of	100	bp
*	51193	55262:	contig of 4070	bp	in length
*	55263	55362:	gap of	100	bp
*	55363	60982:	contig of 5620	bp	in length
*	60983	61082:	gap of	100	bp
*	61083	67628:	contig of 6546	bp	in length
*	67629	67728:	gap of	100	bp
*	67729	72690:	contig of 4962	bp	in length
*	72691	72790:	gap of	100	bp
*	72791	79286:	contig of 6496	bp	in length
*	79287	79386:	gap of	100	bp
*	79387	84351:	contig of 4065	bp	in length
*	84352	83551:	gap of	100	bp
*	83552	92170:	contig of 8619	bp	in length
*	92171	92270:	gap of	100	bp
*	92271	96333:	contig of 4063	bp	in length
*	96334	96433:	gap of	100	bp
*	96434	103218:	contig of 6785	bp	in length
*	103219	103318:	gap of	100	bp
*	103319	112553:	contig of 9233	bp	in length
*	112554	112653:	gap of	100	bp
*	112654	123839:	contig of 10566	bp	in length
*	123240	123339:	gap of	100	bp
*	123340	137921:	contig of 14582	bp	in length
*	137922	138021:	gap of	100	bp
*	138022	157913:	contig of 19682	bp	in length
*	157914	158013:	gap of	100	bp
*	158014	189768:	contig of 31755	bp	in length

FEATURES	
Source	Location/Qualifiers
	1. 189768
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-792C1"
	/clone_lib="RPC1-11 Human Male BAC
	1. 1015
misc.feature	/note="assembly_fragment"
misc.feature	1116. 1284
	/note="assembly_fragment"
misc.feature	2385. 3898
	/note="assembly_fragment"
misc.feature	3999. 5259
	/note="assembly_fragment"
misc.feature	5360. 6912
	/note="assembly_fragment"
misc.feature	7013. 8145
	/note="assembly_fragment"
misc.feature	8245. 9694
	/note="assembly_fragment"
misc.feature	9795. 11004
	/note="assembly_fragment"
misc.feature	11105. 12907

```
misc_feature      /note="assembly_fragment"
13008. .14916
/note="assembly_fragment"
misc_feature      15017. .17114
/note="assembly_fragment"
misc_feature      17215. .18527
/note="assembly_fragment"
misc_feature      18628. .20561
/note="assembly_fragment"
misc_feature      20662. .22484
/note="assembly_fragment"
misc_feature      22585. .25287
/note="assembly_fragment"
misc_feature      25388. .27603
/note="assembly_fragment"
misc_feature      27704. .30020
/note="assembly_fragment"
misc_feature      30121. .33472
/note="assembly_fragment"
misc_feature      33573. .38084
/note="assembly_fragment"
misc_feature      38185. .41069
/note="assembly_fragment"
misc_feature      41170. .45266
/note="assembly_fragment"
misc_feature      45367. .51092
/note="assembly_fragment"
clone_end:r7
vector_side:right"
51193. .55262
/note="assembly_fragment"
55363. .60982
/note="assembly_fragment"
```

Query Match	50.0%	Score 630.8	DB 2	Length 189768
Best Local Similarity	80.0%	Pred. No. 6,7e-191		
Matches 853	Conservative	0	Mismatches 177	Indels 36
			Gaps	2
Qy 352 gaattatggctaattggaggaacaatttggctcattaaaggaagaatttctaagcaga	411			
Db 86596 GACCTTTTAACTTGAAGAAATCATTTTAGGGTATCTGGCAGAAATAATTTCTAAGCAGCA	86537			
Qy 412 aagcatcaagaagtgagcttgggtgtgtctgtctaaagcattcagttctcaaaaggaagcaga	471			
Db 86536 AAGCATTCAGAGGTGACTGGGTGCTGTATTAAGTATTCAGTTTATTAAGGAGACAGCA	86477			
Qy 472 gcataaagatcagaagaatttgcacccctgcacatgtgtataaagaanaaacccatttc	531			
Db 86476 GCATPAAAAGTCTGGAAAAATTTGCAGTCCCAACATGTGTATPAAAAAGAAATTTCAATTTTC	86417			
Qy 532 tgaagggaattcaagcttggtgcagcaaaattgtcattatgtaatgaagagcttgaatttaa	591			
Db 86416 TGAGGAGAAATTCAGCTGGCTGCAGAAATGTGCATTAAGTAAACAGAGAGCTGAAATTTTAA	86357			
Qy 592 tctctcaagacaattggggaaaataatctctctcgtgcacgkctgaagaggtcttcaacagcagtcacat	651			
Db 86356 TCCCAAGACAAAGGGGAAAAACATCTCCAAAGCAGTGCAGAGGTCTTCCAGGGAGCCCT	86297			
Qy 652 caaatcacctggcttggaagccttagagaga---tgittttgttggacagggcccaaggttc	708			
Db 86296 CCCATCTTAAACCCAAAGGCGCTTGGAGCAAAAAGTGTTTGTGGCCAGGCCCCAGGGTTC	86237			
Qy 709 cctgtgtgtgtgtgcagccttagaagaacttgtgtccctgtgtcccaattcaatca-----	760			
Db 86236 CTTGTGCTGTGTGCAGACTTAGGGACTTGGTCCCTGCATCCACACTGCTTCACCCATGCG	86177			
Qy 761 -----gcctgtgcttcaagaaggttgcagagccccaagcctt	795			
Db 86176 TGAAGGGGCCAACATAGACCTCGGGCTGTGGCTTCCAGAAAGTCCAAAGCCCCAAGCTTGG	86117			
Qy 796 gcaagcttcaagtggtgttgagcctgtgtgtgcataaagaatcgaagaattgaatttcggaa	855			
Db 86116 GCACCTTTCATGTGTGTGTGATGCTGTGCAGAGTCCAAAGAGCTCAATAATTTGGGTGGTGGGA	86057			

----- Summary Statistics -----
Contig length: 154090
Phrap values in estimate: 153545
Average error rate (BCM-Phrap estimate): 0.000261755
Fraction of Phrap values less than 40 : 0.0284086
Number of consensus changing edits: 22
Number of N's in consensus : 0

----- Consensus changing edits -----
Position Original+Context Edited+Context
9347 aacacagacc(n)ttttttttt aacacagacc(v)ttttttttt
10135 attgcccct(n)taaggaaaga attgcccct(v)taaggaaaga
10133 ctgtcacacc(n)aaagatcca ctgtcacacc(v)aaagatcca
10417 agcaatgcag(n)ggctacaaga agcaatgcag(v)ggctacaaga
61959 cttactat(n)ttttatct cttactat(v)ttttatct
61960 ttactatn(v)gtttatct ttactatn(v)gtttatct
61961 ttactatn(v)gtttatct ttactatn(v)gtttatct
61962 ttactatn(v)gtttatct ttactatn(v)gtttatct
61983 acaacacaag(n)ntaggttgg acaacacaag(a)ntaggttgg
61993 caacacaag(n)ntaggttgg caacacaag(v)ntaggttgg
61993 nttaggttgg(n)ttactctcc nttaggttgg(v)ttactctcc
62205 tcaatgccc(n)ctgtccatc tcaatgccc(a)ctgtccatc
62956 actgcaacct(n)tgctccag actgcaacct(v)tgctccag
63722 ttacatata(n)caagtaagta ttacatata(v)caagtaagta
90212 cctagaanaa(n)gaactttct cctagaanaa(v)gaactttct
90216 gaaanaagac(n)tttcttctt gaaanaagac(v)tttcttctt
90236 ttcttttta(n)ttttctat ttcttttta(v)ttttctat
90236 taatttttc(n)atcttaagg taatttttc(v)atcttaagg
90443 accttagatg(n)ctctccag accttagatg(a)ctctccag
91405 cacagcttaa(n)caagtaaga cacagcttaa(a)caagtaaga
122280 aaaaagaag(n)gtaaggaaga aaaaagaag(a)gtaaggaaga
137090 acagagaag(n)caaaacacac acagagaag(a)caaaacacac
141632 tgaagccact(n)gaaaagtaac tgaagccact(v)gaaaagtaac

----- Distribution of Quality < 40 Bases -----

#	5	10	15	20	25	30	35	40
10001	*	*	*	*	*	*	*	*
9001	*	*	*	*	*	*	*	*
8001	*	*	*	*	*	*	*	*
7001	*	*	*	*	*	*	*	*
6001	*	*	*	*	*	*	*	*
5001	*	*	*	*	*	*	*	*
4001	*	*	*	*	*	*	*	*
3001	*	*	*	*	*	*	*	*
2001	*	*	*	*	*	*	*	*
1001	*	*	*	*	*	*	*	*
01	*	*	*	*	*	*	*	*

Version: 1.01 qxf.
Location/Qualifiers
source 1..154090
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-13C3"
54..288
repeat_region /rpt_family="MIR"
1020..1146
repeat_region /rpt_family="MER3"
complement(1248..1584)
repeat_region /rpt_family="MER46B"
complement(1624..1682)
repeat_region /rpt_family="MIR"
2630..2761
repeat_region /rpt_family="GA-rich"
4088..4216

repeat_region /rpt_family="L2"
5232..5301
repeat_region /rpt_family="MER102"
complement(5450..5749)
repeat_region /rpt_family="AluY"
6571..6616
repeat_region /rpt_family="(Tc)n"
complement(6678..7320)
repeat_region /rpt_family="LIME3"
7419..7454
repeat_region /rpt_family="AT-rich"
complement(7741..7807)
repeat_region /rpt_family="MER5A"
8113..8267
repeat_region /rpt_family="MER5A"
8385..8587
repeat_region /rpt_family="MIR"
8589..8630
repeat_region /rpt_family="L2"
8666..8695
repeat_region /rpt_family="(TTA)n"

Query Match 49.9% Score 689; DB 9; Length 154090;
Best local similarity 80.1%; Pred. No. 2.2e-190;
Matches 850; Conservative 0; Mismatches 175; Indels 36; Gaps 2;

QY 352 gaataattggtctaagaggaagaacatttggctaaataaggaagaatttctaagcaga 411
DB 101594 GAACTTTGAACTTTGAAACATGATTTAGGTTATCTGCGGAGAAATTTCTTAAGTACA 101653
QY 412 aagcattcaagaggtgacttgggtgctgttaaggaacattcagttcatataggaagcaga 471
DB 101654 AAGCATTAAGAGGAGGACTTGTAGATGTGTTAAAGGTATTCATTAATTAAGGAGGACAGA 101713
QY 472 gcataagagttcagaanaatttgcacccctgacaatgtgataaanaagaaacccatttctc 531
DB 101714 GCATTAAGCTTTGAAATTTTGACGCTGACGCTGACATGATGATGAAANAACCATTTTTC 101773
QY 532 tgaagggaattcaagctgctgcagaaatttgcatatgtaatgaggagctgaaatgttaa 591
DB 101774 TGAAGGGGAAATTTGAAAGCTGCTGCGAATAATTTGCATTAAGTAACAAAGACCTTAATGTAA 101833
QY 592 tctcaagaacaatgggggaaataatctcttgacatgtcagaaggtcttcacagatcat 651
DB 101834 TCCCCAGACAAATGGGGAAATGTCCTCAAGGCTATCAAGAGGCTTCTCATGGACCCCAT 101893
QY 652 caatcacctgacctggaagccttagagaagaatgtttgtggaacagccaggtccct 711
DB 101894 CCCATCACAGGCCCGGAGGCTTAGGAGAAAATGCTTTTGTAGGCCAGGCCCGAGGCCCC 101953
QY 712 gtgctgtgtgcagcttagagactgtgtgcccgtgtt----- 747
DB 101954 ATGCTGTGTGACGCTTAGAGCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 102013
QY 748 -----cccaagttaattcagctgtgtgtctcagaaggtgtcagaagcccaagcttgg 796
DB 102014 GAAAGGCGCCAGCAATAGAGCTTGGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 102073
QY 797 caagcttcaaggtgtgtgtgagcctgtgtgtgcagaagaatgcaagaatgagttggtaa 856
DB 102074 CAATTTCCATGTGTGTGTGTGACCTTGAAGGCTGACAGAGTCAAGAGTCAAGAGTGTGGAG 102133
QY 857 cctcaatcagaattcagaagataatgtaaacccctgtagtccagcagaagtttgtc 916
DB 102134 CCGCCACCGATTTTCAGAGAGTATGAGAAATGCGTGATGTCAGAGAGTGTTCCT 102193
QY 917 gtaagggtgtgggttccctcattgagaacctctgcagaaggtagtlacaaaggaattgtgg 976
DB 102194 TCAAGGGGTGGGCGCTCATGAGAACTCTGTAAGGCAAGTGCAGAGGAATGTGGGG 102253
QY 977 tgggagcccccaacagaagctcccaagtggtgtccatctagtagagctgtggaagaagt 1036
DB 102254 TTGGAGCCCATACACAGAGTCCCTAATGTGGGACCACTTAAGGGGAGCTGTGAGAAAGGG 102313

Oy	1037	ccacatctcttcagactcagaagggtgatctaccatgcagcttgcagtgtgacctga	1096
Oy	1097	aaatcacagacatcagctgcagagccctgtgaagcagcagggatgtagctgttacccctac	1156
Db	102314	CCACATCTCTTCAGACTCAGACCAGAAATGGTAGTGCACCAACAACGCTTGACCATGTACTCGGA	102373
Oy	1097	aaatcacagacatcagctgcagagccctgtgaagcagcagggatgtagctgttacccctac	1156
Db	102374	AAAGCACAGAGACTCAACACACGCCCATGAAAGCAGCCAMAGAGGAGAGCTGTACCTTGC	102433
Oy	1157	aaaaccgtagtggcagagctgcaccaagaacgctgggaattacacctctgtcatgtcatgac	1216
Db	102434	AAAGCCACAGAAGCAGAGACTTTCCCAAAGACACAGGGAACCCACCTCTTGATCATGTTGAC	102493
Oy	1217	ctgagcgtgagagacatgaggtctcaaaagatatcatcttttgagaccttaagatttctacgtcccc	1276
Db	102494	CTAGATGTGAAAACAAGCAGCTCAAAGAACATCATATTAGAACTTTAGATTTGACTACACTT	102553
Oy	1277	actgattctgcgacttatcatatggtggccgcta-ccccttgtttgtggccaatttttccat	1335
Db	102554	TCTGGGTTTTACAGACTTTCGATGGGCGCATRATAGCCCCCTTTGTGTTGGCCAATTTCTCCATT	102613
Oy	1336	tggaaatgcgcgtatttaccatgccaatgctgcttaccctcatgtga	1376
Db	102614	TAGAACAGCTGTATTACTACTCAATGCCGTGATGCCATTGCA	102654
RESULT	9		
AC026107		186660 bp DNA linear HTG_06-FEB-2002	
LOCUS			
DEFINITION	Homo sapiens chromosome 12 clone RP11-307L1, WORKING DRAFT		
VERSION	AC026107		
KEYWORDS	AC026107.22 GI:1389178		
SOURCE	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FUZZTOP.		
ORGANISM	human.		
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
	1 (bases 1 to 186660)		
	Muzny,D.M., Adams,C., Adio-Oduola,B., Alt-osman,F.R., Allen,C.,		
	Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayale,T., Banks,T.,		
	Babarika,J., Benton,J., Bimage,K., Blankenburg,K., Bonin,D.,		
	Bouck,J., Bowe,S., Briviera,M., Brown,E., Brown,M., Bryant,N.P.,		
	Bunyah,P., Burich,P., Burkett,C., Burrell,K.I., Byrd,N.C.,		
	Carron,T.F., Carter,M., Cavazos,S.R., Chaco,J., Chavez,D.,		
	Chen,G., Chen,R., Chen,Z., Chowdry,I., Christopoulos,C.,		
	Cleveland,C.D., Cox,C., Coyle,M.D., Datborne,S.R., David,R.,		
	Davila,M.L., Davis,C., Davy-Carroll,L., Dederlich,D.A.,		
	Delaney,K.R., Delgado,O., Dem,A.L., Ding,Y., Dinh,H.H.,		
	Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,		
	Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,		
	Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,		
	Gadisi,A., Geo,J., Garcia,A., Garner,T., Garza,N., Gill,R.,		
	Gorelli,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,		
	Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,		
	Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,		
	Homsli,F., Howard,S., Huber,J., Huliy.S., Hume,J., Jackson,L.E.,		
	Jacobson,B., Jia,X., Johnson,R., Jolyvet,S., Joudh,S.,		
	Karlsson,E., Kelly,S., Khan,U., Kling,L., Kovach,J., Kovar,C.,		
	Kratovic,J., Kurshige,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,		
	Li,J., Li,Z., Lichtarge,O., Ilean,C., Liu,Q., Liu,W., Louisedge,H.,		
	Lodado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,		
	Maneshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,		
	Massey,E., Mcleody,E., Mcloed,M.P., Meador,M., Mel.G., Metker,M.,		
	Miner,G., Miner.Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,		
	Moser,M., Neal,D., Newcom,J., Newton,S., Nguyen,A., Nguyen,N.,		
	Nguyen,N., Nickerson,E., Nwokwenko,S., Oguh,M., Okwuonu,G.,		
	Ongunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,		
	Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,		
	Rives,M., Rojas,A., Rojnbokan,I., Rolfe,M., Ruiz,S., Savory,C.,		
	Scherer,S., Scott,G., Shen,H., Shoostrati,N., Sisson,I.,		
	Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,		
	Stuton,A., Svatek,A., Tabor,P., Tamerisa,A., Tameisla,K., Tang,H.,		

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Unpublished
2 (bases 1 to 186660)
Worley,K.C.
Direct Submission
Submitted (19-MAR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 1, 2001 this sequence version replaced gi:1387175.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: HAMK
Center clone name: RP11-307L1
----- Summary Statistics
Sequencing vector: Plasmid;
Sequencing vector: M13;
Chemistry: Dye-primer Bodipy: 18% of reads
Chemistry: Dye-terminator Big Dye: 82% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 186630 bases at least Q40
Consensus quality: 186659 bases at least Q30
Consensus quality: 186660 bases at least Q20
Estimated insert size: 186779; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-1p estimation
Quality coverage: 10.7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 186660: contig of 186660 bp in length.
Location/Qualifiers
1. 186660
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-307L1"
BASE COUNT 57089 a 37346 c 35962 g 56263 t
ORIGIN

Query Match 49.9%; Score 689; DB 2; Length 186660;
Best Local Similarity 80.1%; Pred. No.2,3e-190; Indels 36; Gaps 2;
Matches 850; Conservative 0; Mismatched 175;

Db 352 gaataatggcctcaaggaagcaaatcttggtctataagaagaattctaaacga 411
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 71843 GAACCTTGAACTTGAGAGACATGATTAGGCGTATCGCGGGAAGAAATTTCTAAGTACGA 71902

QY 412 aagcatcaagaagtgactggggtgcgtctaaagcattcagttcatcaaggagcaga 471
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 71903 AAGCATTTAAAGAGCGTGAATGATGTGCTTAAAGCTATTTCATTATTAAGGAGCGCAGA 71962

QY 472 gataatggctcaagaatttgccacctgcgaacatctgtttaaaaaaagaaacccatttc 531
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 71963 GCATTAAGCGTTTGGAAATTTGCGACCGCTGCACATGTGATAGGAAAGAAACCCATTTC 72022

```


Db	25294	TTGGTAACCTCTGCCTACATATTTCAAGAGTGATGAAACCCCTGGATGCCAACGAAA	25235
Qy	910	gtttctgtatgaggtgtggtgtctctcatlgtagaagaccctctcaaggtgtatgtacaaagggaaa	969
Db	25234	GTCTCTCAGAGGGGTGGGGCCCTCTATGGAGAACTCTCTGTAGGGCAGTGTCAAGGGAAA	25175
Qy	970	tgtttgtgtgtgagccccacacagagctcccccagctgtgtgtccatctagttagagctgtgag	1029
Db	25174	TGTGGGGTTGAAGAGCCCCACACAGAGTGTCCCTACTTGTGGGACCTCTAGTGGAGCTGTGAG	25115
Qy	1030	aagaagtcacacatctctccagaactccagaaggtatgatctcactgtcacagctgtcagatgt	1089
Db	25114	AAGAGGGCCACCACTCCAGACCCACAGATMGTTGATTCACCTGACAACTTGCACCTGTGA	25055
Qy	1090	gacctaaatccacagacactcagctgtccagctgtgaagaagacagagatgtgagctgt	1149
Db	25054	GCGTGGAAAAGCCACAGACACTCAACCGTAGCTTGTGAAGACAGCCAGAGGAGAGCTCT	24995
Qy	1150	accctacaaaacgtagtgtgcagagctgcacaaagacgctgtggaatctacctcttgcattg	1209
Db	24994	ACCCTGCAAAAGCCACAGGGGCGATACCTGCCCAAGACATGGGAACCCACCTTGATCA	24935
Qy	1210	tcatgacctgtgacgtgagacactgtggtctaaagaatatctttgtagcttgaatttga	1269
Db	24934	GTGTACCTGTCATATGAACGTGGAGTCAAAAAGATCATTTTGGATCTTTAAGACTTGA	24875
Qy	1270	ctgcccacactgtgattcttgagactatattgtggggccgcta-cccccttgttttggccaattt	1328
Db	24874	CTGTCTACTGTGATTTACAGACTCCGATGGGGCGTGTAGCCCTTTGTTTGGCCAAATGTC	24815
Qy	1329	ttccatttgggaactgcggtatttaccacaatgcctgtacctcatgtatgtgag	1381
Db	24814	TCCCATTTAGATGAGCTGTATTATTATCAATACCTGTACCCCATTTGTATCTGAG	24762
RESULT	11		
AC010312			
LOCUS	AC010312	67984 bp	DNA linear PRI 23-AUG-2001
DEFINITION	Homo sapiens chromosome 5 clone CTR-62P13, complete sequence.		
ACCESSION	AC010312		
VERSION	AC010312.5	GI:15281196	
KEYWORDS	HTG.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryote; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	DOE Joint Genome Institute and Stanford Human Genome Center.		
JOURNAL	Direct Submission		
REFERENCE	2 (bases 1 to 67984)		
AUTHORS	DOE Joint Genome Institute.		
TITLE	Direct Submission		
JOURNAL	Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint		
REFERENCE	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA		
AUTHORS	3 (bases 1 to 67984)		
TITLE	DOE Joint Genome Institute and Stanford Human Genome Center.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (03-NOV-2000) DOE Joint Genome Institute, 2800 Mitchell		
AUTHORS	4 (bases 1 to 67984)		
TITLE	Drive, Walnut Creek, CA 94598, USA		
JOURNAL	Direct Submission		
REFERENCE	Submitted (26-NOV-2000) DOE Joint Genome Institute, 2800 Mitchell		
AUTHORS	5 (bases 1 to 67984)		
TITLE	Drive, Walnut Creek, CA 94598, USA		
JOURNAL	Direct Submission		
REFERENCE	Submitted (23-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell		
AUTHORS	Drive, Walnut Creek, CA 94598, USA		
TITLE	On Aug 23, 2001 this sequence version replaced gi:11079410.		
JOURNAL	Draft Sequence Produced by DOE Joint Genome Institute		
COMMENT	www.jgi.doe.gov		

Finishing Completed at Stanford Human Genome Center									
www.sngc.stanford.edu									
Quality: Phrap Quality >40 99.7% Of Sequence:									
Estimated Total Number of Errors is 0.1.									
STS Content:									
SHGC-10855 G13672.									
Location/Qualifiers									
1. .67984									
/organism="Homo sapiens"									
/db_xref="taxon:9606"									
/chromosome="15"									
/clone="CTB-62P13"									
BASE COUNT 19168 a 13195 c 14662 g 20959 t									
ORIGIN									
Query Match	49.7%;	Score 686;	DB 9;	Length 67984;					
Best Local Similarity	81.9%;	Pred. No. 1.6e-189;							
Matches 830;	Conservative	0;	Mismatches 150;	Indels 34;	Gaps 2				
Oy	402	ctaaagcagaacagcattcaagagcgcagctgggtgctgtcttaagagcattcagttcataa	461						
Db	7119	ctaaacagcaaaagcattctcaagaaagttatgtgtgtctgttaaaagcactcggtttatata	7178						
Oy	462	ggagagcgagcagataagagttcagaanaatttgcacccctgcacaaatgtaataaagaanaa	521						
Db	7119	gggaagcagagcatttgaattttggaaaatttttcagcctcagcctacatgctgtagaaaagaaaa	7238						
Oy	522	accacatttcctgaggggaaatcaaatcagctgctgcagcagaanaatttgcataatattagagac	581						
Db	7239	accattttctgagaggaatgaattgaagccagctgagaaattttgcatttaattatagagagac	7298						
Oy	582	tgaattgaatccctcaagaacaaatggggaaataatactcctctgcagacatgctcagagttctcac	641						
Db	7289	caaatgttaattcccccaagacaaatggggaaaattgtctccggagacatgacagagttttcac	7358						
Oy	642	agcagttcatcaatcaatcattcagcttgccttgagggccttagagaanaattggttttggagacagcc	701						
Db	7359	agcaccactccccaacacagcagcccaaaagcctttggagaaaatcgtttctggagcccaagcc	7418						
Oy	702	cagagttccctgtgctgtgtcagccttagagacttggtgtgcctgtgtccagtttaattcac	761						
Db	7419	cagagttccctgtgctgtgtcagccttagagacttggtgtgcctgtgtccagtttaattcac	7478						
Oy	762	c-----tggtgcttcagaagggtgcagccca	788						
Db	7479	ccatgacataaaaaagtgcccaaaactacatttcagagttgtttcttcagaaagtggaagccca	7538						
Oy	789	agccttgacagcttcacaagtgtgtctgaagcctgtggtgccaagaagtcacaagaattgag	848						
Db	7539	agccttgacagcttcacaagtgtgtgtgagcctgtgagcctgcagcagcaagaaagtcacaagaattgag	7598						
Oy	849	tttggaacctccaatcagatttcagaagaataatgaaaccccttgatgcccagagcaga	908						
Db	7599	tttggttaacctctgctagatttcagaagaataatgaaaccccttgatgcccagagcaga	7658						
Oy	909	agtttgcgtgtgaggtgtggtgtcctcatatggagaacctctgcagaaggtatgatacaaaaggaa	968						
Db	7659	agtttgcgtgtgaggtgtggtgtcctcatatggagaacctctgcagaaggtatgatacaaaaggaa	7718						
Oy	969	atgtgtgtgtggagagccccaacagaagtcocccagctggaggtcctacactagtagagagctgta	1028						
Db	7719	atgtgtgtgtggagagccccaacagaagtcocccagctggaggtcctacactagtagagagctgta	7778						
Oy	1029	gaagaagtcacacacatcctccagactccagaagcaggtatagatcccatgacagcttgcagcat	1088						
Db	7779	gaagaagtcacacacatcctccagactccagaagcaggtatagatcccatgacagcttgcagcat	7838						
Oy	1089	tgcttgaaaaatccaacagacactcagctgcagagccctgtgaagcagcagagatgagagctcg	1148						
Db	7839	agccttgaaaaatccaacagacactcagctgcagagccctgtgaagcagcagagatgagagctcg	7898						
Oy	1149	tacctataaaaacgtatgtgcagagctgcagcaaaagcgttgagatcttacccttgatc	1208						

Db	7959	TACCGTCAAAAGCACACAGGGGCGATAGCTGCCCAAGACCATGGGAACCCACCCCTTGATC	7958
QY	1209	gtcatgcctgtagcgctgtagacatgagtcacaaagatcatlcttgtagcttlaagattg	1268
Db	7959	AGTGTGACCTGCATATGACAGACGTGGAGTCACAAAGAGATCATTTGGATCTTTAAAGCTTG	8018
QY	1269	actgccccatgattcttgtagactatataggggcccgta-ccccctgttcttgcccaatt	1327
Db	8019	ACTTCTCTACTGTGATTTTCAGACTCGCATGTGGGCGCTGTAGGCCCTTTTGTGGCCAAATGT	8078
QY	1328	ttccacttggaactgcgcgtacttaaccatgacgcgtacccctcatgtatgtag	1381
Db	8079	CTCCCATTTAGATAGCTGTATTTATTCACATACCTGTACCCCACTTGTATCTAG	8132
RESULT	12		
AC091005/c			
LOCUS			
DEFINITION	AC091005	168502 bp	DNA linear prt 11-DEC-2001
ACCESSION	AC091005	Homo sapiens chromosome 15, clone RP11-1008C21, complete sequence.	
VERSION	AC091005.9	GI:17488655	
KEYWORDS		HTG.	
SOURCE		human.	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Carnivora; Homnidae; Homo.	
AUTHORS		1 (bases 1 to 168502)	
TITLE		Biiren,B., Linton,L., Nusbaum,C. and Lander,E.	
JOURNAL		Homo sapiens chromosome 15, clone RP11-1008C21	
AUTHORS		Unpublished	
		2 (bases 1 to 168502)	
		Biiren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S., Barne,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A., Camarata,J., Campoliano,A., Chang,J., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Deatrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzhugh,N., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heatford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Karlitas,A., LaRoque,K., Lamaizars,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C., Macdonald,P., Margulis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Mcpheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Plerre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rossi,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Sougnuez,C., Spencer,B., Strange,Thomann,N., Stojanovic,N., Strauss,N., Sudirmanian,A., Talamas,J., Tesfaye,S., Theodore,J., Travers,M., Travis,N., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and zody.M.	
TITLE		Direct Submission	
JOURNAL		Submitted (24-MAR-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	
REFERENCE		3 (bases 1 to 168502)	
AUTHORS		Biiren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barne,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A., Camarata,J., Campoliano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., Deatrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzhugh,N., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goid,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heatford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamt,A., Karatas,A., Kells,C., Labocque,K., Lamaizars,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C., Macdonald,P., Major,J., Margus,N., Matthews,C., McCarthy,M., McKernan,K., Mcpheeters,R., Meldrim,J., Meneus,L., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Plerre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Sougnuez,C., Spencer,B., Strange,Thomann,N., Stojanovic,N., Strauss,N., Sudirmanian,A., Talamas,J., Tesfaye,S., Theodore,J., Travers,M., Travis,N., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and zody.M.	

[illegible]

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE
AUTHORS 1 (bases 1 to 131215)
TITLE Sulston, J.E. and Waterston, R.
JOURNAL Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)

MEDLINE
99063792

REFERENCE
AUTHORS 2 (bases 1 to 131215)
TITLE Nguyen, C., Abbott, A. and Elliott, G.
JOURNAL The sequence of Homo sapiens BAC clone RP11-42414
Unpublished (2001)

REFERENCE
AUTHORS 3 (bases 1 to 131215)
TITLE Waterston, R.H.
JOURNAL Direct Submission
Submitted (04-SEP-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108 USA

REFERENCE
AUTHORS 4 (bases 1 to 131215)
TITLE Waterston, R.H.
JOURNAL Direct Submission
Submitted (05-SEP-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108 USA

REFERENCE
AUTHORS 5 (bases 1 to 131215)
TITLE Waterston, R.
JOURNAL Direct Submission
Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Sep 5, 2001 this sequence version replaced g1:15209227.

COMMENT

Genome Center

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.wustl.edu

Summary Statistics

Center project name: H_NH0424104

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, E.,
Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(<http://bacpac.med.buffalo.edu>)
VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-22119, 2000 bp overlap; the
clone sequenced to the right is RP11-364115. Actual start of this
clone is at base position 128235 of RP11-22119; actual end is at
base position 131215 of RP11-42414.

FEATURES
source
The sequence between 58007 and 58074 is covered only by a PCR
product of clone DNA; the sequence between 58180 and 58221 is
covered only by a PCR product of clone DNA. Data from AC093650 was
used to finish this clone, AC079614. Polymorphisms exist between
RP11-42414 and RP11-364115.
Location/Qualifiers
1..131215
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-42414"
/clone_id="RP11-11"
1..298
/rpt_family="L1"
repeat_region
299..594
/rpt_family="Alu"
595..805
/rpt_family="L1"
873..916
/rpt_family="MaLR"
917..953
/rpt_family="ERV_L"
954..1260
/rpt_family="Alu"
1236..1260
/rpt_family="(A)n"
1261..1751
/rpt_family="ERV_L"
1752..1995
/rpt_family="MaLR"
1996..2307
/rpt_family="MaLR"
2308..2609
/rpt_family="Alu"
2610..2817
/rpt_family="MaLR"
2819..2911
/rpt_family="MaLR"
2971..3072
/rpt_family="MERL_type"
3143..3270
/rpt_family="(CATARA)n"
3361..3427
/rpt_family="MERL_type"
3453..3706
/rpt_family="MERL_type"
5132..5339
/rpt_family="MIR"
5483..5702
/rpt_family="MERL_type"
6204..6274
/rpt_family="MERL_type"
6914..7012
/rpt_family="AT_rich"
7467..7498
/rpt_family="AT_rich"
8317..8542
/note="match to EST AW899791 (MID:98063996) "
8908..8923
/note="match to EST BT492884 (MID:91332228) "
8908..8938
/rpt_family="(T)n"
8909..9218
/rpt_family="Alu"
9227..9255
/rpt_family="AT_rich"
10278..10489
/rpt_family="MIR"
10647..10943
/rpt_family="Alu"
11723..11855

DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE HP-20.
GN HP-20.
OS *Tamias sibiricus* (Siberian chipmunk).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae;
OC *Tamias*.
OC NCBI_TaxID=64680;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2188336; PubMed=11602349;
RA Oho M., Hosoe Y., Azuma S., Shoji M., Nara K., Kondo N., Shiba T.,
RA Takamatsu N.,
RT "HNF-1 regulates the liver-specific transcription of the chipmunk HP-
RT 20 gene.";
RL Gene 277:121-127(2001).
DB EMBL: AB067779; BMB6362.1; -
DR SEQUENCE 196 AA; 21330 MW; B07D17BFA94D2FD CRC64;

	Query Match	13.0%	Score 324;	DB 11;	Length 196;
	Best Local Similarity	46.3%:	Pred. No. 1,9e-21;		
	Matches 68;	Conservative 21;	Mismatches 46;	Indels 12;	Gaps 4;
Oy	1 IVIVPLVITRVIEHVNAGP---AHPRPPEVG-----PGAPGLPYQTGISEMTC	51			
	: : : : : : : :				
Dd	7 LAIVLVNVLNVNQVSCSGPPGVETPCGVPPGRPGQPGAAGRPGDPRKGSYVC	66			
Oy	52 PCPDIESAFVTLGSKLPLEPKPIITFGVLVYNMQRDLEAMGVACRVGNYSFDFVE	111			
	: : : : : : : :				
Dd	67 PCR--ESSAFTVFNSRLPPSEPVTFTEVLVYNMQDLKASTGVENCVEGHNHFSDVE	124			
Oy	112 LHCKVNIWIMRKOT-LANKEELISQQ	137			
	: : : : :				
Dd	125 LYHCVRIGLIMKNHIQMEKHQLSSNE	151			

Q9H720	PRELIMINARY:	PRT:	222 AA.
ID	Q9H720		
AC	Q9H720;		
DT	01-MAR-2001 (TrEMBLrel. 16, Created)		
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)		
DT	01-OCT-2001 (TrEMBLrel. 18, Last annotation update)		
DE	HYPOTHEITICAL 24.1 KDA PROTEIN.		
OS	Homio sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
OX	NCBI_Taxid=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=EMBRYO;		
RA	Isoegal T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,		
RA	Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,		
RA	Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,		
RA	Arita M., Nebekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,		
RA	Makatsutsu A., Nakamura Y., Nagahari K., Masuno Y., Oshima A.,		
RT	"NEDD human cDNA sequencing project."		
RL	Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.		
SR	EMBL; AK024120; BAB1833.1;		
SO	SEQUENCE 222 AA; 24079 MW; 646ED514F9F541D9 CRC64;		

```

Query Match 11.0%; Score 275; DB 4; Length 22;
Best Local Similarity 40.0%; Pred. No. 6, 1e-17;
Matches 76; Conservative 22; Mismatches 64; Indels 26; Gaps
7;

Oy 223 RPR----RKWFGCTGFGS--LCCVDP-----RDLYPCPVNNAVASGASPKFMQLPSGV 271
    |||      |||      |||      |||      |||      |||      |||      |||
Db 26 RPRGLGGRDGFMDQALGSPALCRLLTGWYPAQRKLOPC--LKEAIVQVG--WLRVVEV 79
    |||      |||      |||      |||      |||      |||      |||      |||
Ox 272 EPVGA-----KKSRIEWEPPPIRFQKLYGNPNMPRQKFAVNGSGSSMRTSARVYQR 321
    |||      |||      |||      |||      |||      |||      |||      |||

```

[illegible]

RESULT	4
ID	09Z1K4
AC	09Z1K4
DT	01-MAY-1999 (TREMBlrel. 10, Created)
DT	01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT	01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE	COLLAGEN ALPHA 1 TYPE X (FRAGMENT).
GN	COL10A1.
OS	Rattus norvegicus (Rat).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus
OX	NCB1,TaxID=10116;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=SPRAGUE-DAWLEY;
RX	MEDLINE=40310874; PubMed=10853827;
RA	Marks S.A., Lundmark C., Christersson C., Wurtz T., Odgren P.R.,
RA	Seifert M.F., Mackay C.A., Mason-Savay A., Popoff S.E.;
RT	"Endochondral bone formation in toothless (osteopetrotic) rats:
RT	failures of chondrocyte patterning and type x collagen expression.";
RL	Int. J. Dev. Biol. 44:309-316(2000).
DR	EMBL; A0131848; CAA10518.1; -.
DR	InterPro; IPR001073; Clq.
DR	InterPro; IPR000087; Collagen.
DR	Pfam; PF00386; Clq; 1.
DR	Pfam; PF01391; Collagen; 2.
DR	PRINTS; PRO0007; COMPLEMENTC1Q.
DR	SMART; SM00110; Clq; 1.
DR	PROSITE; PS01113; Clq; 1.
FT	NON_TER
SEQUENCE	295 AA; 30012 MW; FF43B1548028813E CRC64;

	Query Match	6.8%;	Score 169;	DB 11;	Length 295;
	Best Local Similarity	28.4%;	Pred. No. 3.7e-07;		
	Matches 54;	Conservative 19;	Mismatches 69;	Indels 48;	Gaps 6;
Oy	18 AGPAHPRPDEEVGPPQCAPCLPOLYTGSEIEMKCCPD-----	55			
Db	101 AGPPRGRCPPGHGSEGLPPCPPPPCPSPQAVIPPGFTRKSGQRRLSGMPLVSAING	160			
Oy	56 ---IERSAFYVILSGKLPLPFKPIIFTGVLYNAQDLKEMGVFACRVPGNYSSPDVEL	112			
Db	161 VTGMPVSAFTVILSKAPVANGAIPDEILYNRQOHYPBSRGIFTCKIPGIYFYSYHIH	220			
Oy	113 H--HCKYINIMLMKQILLANKELSK-----QSDIQDVT-----WVLKAFSFIKREAH	158			
Db	221 KGTIVHWGLKLNKCTPTMYTYDEYSKYGLDQASGSAIMELTENDDVWKLQP-----NAES	274			
Oy	159 K---SSENLH 165				
Db	275 NGLYSSEYVH 284				
RESULT	5				
096D07					
ID	096D07	PRELIMINARY;	PRT;	744 AA.	
AC	096D07;				
DT	01-DEC-2001 (Tremblrel, 19, Created)				
DT	01-DEC-2001 (Tremblrel, 19, Last sequence update)				

DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HYPOTHEICAL 73.4 KDA PROTEIN.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LUNG CARCINOMA;
 RA Strausberg R.;
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC013581; AAI3581.1;
 KW Hypothetical protein.
 SO SEQUENCE 744 AA; 73364 MW; 2BC1B0955DEC9A3 CRC64;

Query Match 6.5%; Score 162.5; DB 4; Length 744;
 Best Local Similarity 30.4%; Pred. No. 5.1e-06;
 Matches 42; Conservative 17; Mismatches 50; Indels 29; Gaps 5;

OY 19 GPPAHPPEEYGPAGBL-----PQYTGEL-----SEMTKCPDPI 56
 DB 554 GPPGLPGRPPGPPGPPVAVMPPTPOGEVLPDMGLGIDGVKPPHAGAKKNGGPPAY 613
 OY 57 ERSATVRLSGKLPFPPIITGVLYNAQRDLKAMGVACRVGNYSFDELHCK 116
 DB 614 EMPATTAELTAPFPVPGAPVFNKLLYNGRONYNPOTGFTCEVGVYFAVAV---HCK 670
 OY 117 -VNIMLRKOILANKEEI 133
 DB 671 GGNVAV---ALFKNEPV 685

RESULT 6
 ID 096NP0 PRELIMINARY; PRT: 216 AA.
 AC 096NP0:
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CDNA FLJ30449 FIS, CLONE BRACE2009274.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=CEREBELLUM;
 RA Nilmolaya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
 RA Furuuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
 RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
 RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
 RA Suzuki Y., Sugano S., Nagahari K., Masuno Y., Nagai K., Isogai T.,
 RL "NEDO human cDNA sequencing project."
 RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK055011; BAB70839.1;
 SO SEQUENCE 216 AA; 23140 MW; 976A49B029A84D36 CRC64;

Query Match 6.3%; Score 158; DB 4; Length 216;
 Best Local Similarity 28.1%; Pred. No. 2.4e-06;
 Matches 59; Conservative 19; Mismatches 68; Indels 64; Gaps 7;
 OY 209 CORSSQSIKSLAMP-----RRKMPG--TGPGSLCCV-----QPRDLVPCVPVNSAV 255
 DB 15 CORSSM---LPLOYRPPGRLRGKNGMGOSQSPANCILITGWCPSAQIQLQMLMLKGAN 71
 OY 256 ASEGASPKPMQLPSGVEPVGAKKSRLEIV-----EPPI 288
 DB 72 VOLG-----PWLDORVATSLGFGHMYGLMVKRORRRFGSLHLDRGCMEMPGCCGRTPL 127

OY 289 RFQKIYGNPMPROKFRVAVGVSSWRTSARVQKGNVMEPPRRHVSAPSSAARRSPS 348
 DB 128 QGQSPHGDPLGQ-----CRGNGVLEPSORSVSTRALPANGAVGRPSS 170
 OY 349 SRLQGRSTDSQHYPEKSTDPCCOPVKA 378
 DB 171 SRPHKGGSTDSLHCVPKAVGTCOPIRKA 200

RESULT 7
 ID 09D2V4 PRELIMINARY; PRT: 744 AA.
 AC 09D2V4:
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE PROCOLLAGEN, TYPE VIII, ALPHA 1.
 GN COL8A1.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=KIDNEY;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamamata R.,
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikiado I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bona M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gwinnstein S., Hill D., Hofmann M., Hume D.A., Kanuya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seiya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Wetz C., Whitaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL: AK018742; BAB31383.1;
 DR MGI: 88463; Col8a1.
 DR InterPro: IPR001073; C1q.
 DR InterPro: IPR000087; Collagen.
 DR Pfam: PF00386; C1q; 1.
 DR Pfam: PF01391; Collagen; 7.
 DR PRINTS: PRO0007; COMPLEMENTC1Q.
 DR SMART: SM00110; C1Q; 1.
 DR PROSITE: PS01113; C1Q; 1.
 SO SEQUENCE 744 AA; 73581 MW; C659BDCBCBDEB9C CRC64;

Query Match 6.3%; Score 157.5; DB 11; Length 744;
 Best Local Similarity 32.0%; Pred. No. 1.4e-05;
 Matches 40; Conservative 14; Mismatches 46; Indels 25; Gaps 4;

OY 19 GPPAHPPEEYGPAGBL-----PQYTGEL-----ISEMTKCPDPI 57
 DB 555 GPPGLPGRPPGPPGPPVAVMPPTPOGEVLPDMGLGIDGVKPPHAYAGAKKNGGPPAY 614
 OY 58 RSAFTVRLSGKLPFPPIITGVLYNAQRDLKAMGVACRVGNYSFDELHCK- 116
 DB 615 MPATTAELTAPFPVPGAPVFNKLLYNGRONYNPOTGFTCEVGVYFAVAV---HCKG 671
 OY 117 VNIMLRKOILANKEEI 121

Db 672 GNVWV 676

RESULT 8

0921S8 PRELIMINARY: PRT: 744 AA.

AC 0921S8; 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DE PROCOLLAGEN, TYPE VIII, ALPHA 1.

OS Mus musculus (mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI_TaxID=10090;

RP SEQUENCE FROM N.A.

RA Strausberg R.

RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC011061; AA011061.1; -.

KW COLLAGEN.

SQ SEQUENCE 744 AA; 73621 MM; BF7A7FD79D8463AA CRC64;

Query Match 6.3%; Score 157.5; DB 11; Length 744;

Best Local Similarity 32.0%; Pred. No. 1.4e-05;

Matches 40; Conservative 14; Mismatches 46; Indels 25; Gaps 4;

Qy 19 GPPAHPPEEVEGPPAPGL---PQYTCG-----ISEMTKCPDIE 57

Db 555 GQPGDLGPPGPPGPPPPVMTPTSPGCEYLPDMGLGIDGKPPHAKGKKGHPAYE 614

Qy 58 RSAFTYKLSGKLPKPKPIFTGVLYNAORDKEAMGVACRPVGNYSFDELHCK- 116

Db 615 MPAFTELTVPEPPVGPAPKFDKLLYNGRONVNPQTGFTECVGVYFAVHY---HCKG 671

Qy 117 VNIML 121

Db 672 GNVWV 676

RESULT 9

09N178 PRELIMINARY: PRT: 675 AA.

AC 09N178; 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DE TYPE X COLLAGEN.

GN COL10A1.

OS Sus scrofa (pig).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.

NCBI_TaxID=9823;

RP SEQUENCE FROM N.A.

RX MEDLINE=21015405; PubMed=11130976;

RA Nielsen V.H., Bendixen C., Arnbjerg J., Sorensen C.M., Jensen H.E., Shukri N.M., Thomsen B.;

RT "Abnormal growth plate function in pigs carrying a dominant mutation in type X collagen."

RL Mamm. Genome 11:1087-1092(2000).

DR EMBL; AF222861; AAF3271.1; -.

DR InterPro: IPR001073; Clq.

DR InterPro: IPR000087; Collagen.

DR Pfam: PF00386; Clq; 1.

DR PRINTS; PR00007; COMPLEMENTC1Q.

DR SMART; SM00110; Clq; 1.

DR PROSITE; PS01113; Clq; 1.

KW COLLAGEN.

SQ SEQUENCE 675 AA; 65447 MM; 26397B10310383P9 CRC64;

Query Match 6.3%; Score 157; DB 6; Length 675;

Best Local Similarity 32.1%; Pred. No. 1.4e-05;

Matches 45; Conservative 14; Mismatches 53; Indels 28; Gaps 5;

Qy 19 GPPAHPPEEVEGPPAPGLPQYTCG-----GEISE-----MTKCPDIE 57

Db 491 GPPGPPGPKHAGEPGLPDPGPPGPPGQAVPEGEFVKEGQAFVSANQVTMPV---- 546

Qy 58 RSAFTYKLSGKLPKPKPIFTGVLYNAORDKEAMGVACRPVGNYSFDELH--HC 115

Db 547 -SAFTYLSKAYATGAPAPFDKLLYNGQHYDPKVTGFTCRIGITYFSYHVKGTHA 605

Qy 116 KVNIMLRKQILANKKEISK 135

Db 606 WVGLYKNGTPVMTYDEYVK 625

RESULT 10

095J95 PRELIMINARY: PRT: 194 AA.

AC 095J95; 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DE ADIPOLECTIN (FRAGMENT).

GN APM1.

OS Canis familiaris (dog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

NCBI_TaxID=9615;

RP SEQUENCE FROM N.A.

RC TISSUE=ADIPOSE TISSUE;

RA Kabir M., Anantharayan S., Ionut V., Kim S.P., Van Clitters G.W., Dea M.K., Bergman R.N.;

RT "Regulation of Adiponectin gene expression in the fat-fed dog."

RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF417206; AAL09702.1; -.

FT NON_TER 1 194

FT NON_TER 1 194

SQ SEQUENCE 194 AA; 20890 MM; 3AA3D947D187AF9A CRC64;

Query Match 6.2%; Score 155.5; DB 6; Length 194;

Best Local Similarity 34.5%; Pred. No. 3.5e-06;

Matches 41; Conservative 16; Mismatches 55; Indels 7; Gaps 2;

Qy 31 GPPGAPGLPQYTCGEISEMTKCPDIEISATPVKLSGKLPKPKPIFTGVLYNAORDLK 90

Db 67 GPPGFPETPERKGEPEESAY----VHRSFVSGLESHITVNPVPIRFTKIFYNLQNHXD 121

Qy 91 EAMGVACRPVGNYSF--DVLEHCKVNIMLRKQILANKKEISKQCSIOEVTWLL 147

Db 122 GTCKFRHCNIPGLYFSYHITVLYLKDVKVSLYKKKAKLFTYDQYQKNVDQSGSVLL 180

RESULT 11

095M04 PRELIMINARY: PRT: 240 AA.

AC 095M04; 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DE ADIPOSE TISSUE-SPECIFIC PROTEIN ADIPO Q.

OS Bos taurus (bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecorida; Bovidae; Bovinae; Bos.

NCBI_TaxID=9913;

RP SEQUENCE FROM N.A.

RX MEDLINE=21369933; PubMed=11382781;

RA Sato C., Yasukawa Z., Honda N., Matsuda T., Kitajima K.;
RT "Identification and Adipocyte Differentiation-dependent Expression of
RT the Unique Distal Acid Residue in an Adipose Tissue-specific
RT Glycoprotein, Adipo O.,"
RL J. Biol. Chem. 276:28849-28856(2001).
DR EMBL: AF269230; AAK58902.1; -.
SQ SEQUENCE 240 AA; 26091 MM; C6253BA803B9A668 CRC64;

Query Match 6.2%; Score 155.5; DB 6; Length 240;
Best Local Similarity 29.7%; Pred. No. 4.7e-06;
Matches 43; Conservative 27; Mismatches 64; Indels 11; Gaps 3;

OY 31 GPPGAPGLPQYTGISEMTKPCPDIERSAFTVLSGLPLPFPKPIITFTGLVNAQRDLK 90
DB 85 GPRGPPGPRGRGEPGEALAY-----VYSAFSGVLETRTVNVPNFRKTIKTYNQNHVD 139
OY 91 EAMGVFACRVGNVYSSDVELH--HCVNITLMLRKQILANKKEISQOSIOEVTWVLK 148
DB 140 GSTGKFCYNIPGLYFSYHITYMKDVKYSLEKDKKAVLEFYDOYQEKNVDAQSGSVLLH 199
OY 149 A----FSEFREAHEKSSENLHPDNV 169
DB 200 LEVGDQVWLQYVEGENHNGVYADNV 224

RESULT 12
OY0MB3 PRELIMINARY; PRT; 350 AA.

AC OY0MB3.
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MYELOID-SPECIFIC PEROXIDASE (FRAGMENT).
GN MPX.
OS Brachdano reio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Danio.
RX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-KIDNEY;
RA Lleschke G.J., Oates A.C., Crowhurst M.O., Ward A.C., Layton J.E.;
RT "Macrophages in embryonic and adult zebrafish."
RL Blood 0:0-0(2001).
DR EMBL: AF378825; AAK91662.1; -.
KW Peroxidase.
FT NON_TER 1
SQ SEQUENCE 350 AA; 38647 MM; 4E1EADDA0573F70 CRC64;

Query Match 6.0%; Score 148.5; DB 13; Length 350;
Best Local Similarity 33.3%; Pred. No. 3.4e-05;
Matches 34; Conservative 12; Mismatches 33; Indels 23; Gaps 3;

OY 17 VAGPPAHPRPPEE-----VGPAPGLP---QYTGISEMTKPCPDIERSAFTVLS 66
DB 244 IPGPPGPRGPRGERGPRGVAGPRGPRGIPRPINTG-----QSSAFASVNV 290
OY 67 GKLPPLFPPIITFTGLVNAQRDLK EAMGVFACRVGNVYSSF 108
DB 291 SILPATAKVVVGVLYNGONHNYNOTSGMFLQIPGVYEFEP 332

RESULT 13
OYBXJ2 PRELIMINARY; PRT; 289 AA.
AC OYBXJ2.
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)

DE COMPLEMENT-C1Q TUMOR NECROSIS FACTOR-RELATED PROTEIN.
GN CTRP7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Piddington C.S., Sheppard P.O., Bishop P., Lasser G.W.;
RT "Homo sapiens complement-c1q tumor necrosis factor-related protein."
RT Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF329839; AAK17963.1; -.
DR InterPro: IPR001073; C1q.
DR InterPro: IPR000087; Collagen.
DR Pfam: PF00386; C1q; 1.
DR Pfam: PF01391; Collagen; 1.
DR PRINTS: PR00007; COMPLEMENTC1Q.
DR SMART: SM00110; C1Q; 1.
DR PROSITE: PS01113; C1Q; 1.
SQ SEQUENCE 289 AA; 30683 MM; A61609FF86D26946 CRC64;

Query Match 5.8%; Score 144; DB 4; Length 289;
Best Local Similarity 36.2%; Pred. No. 6.7e-05;
Matches 38; Conservative 11; Mismatches 46; Indels 10; Gaps 2;

OY 19 GPPAHPRPPEEYGPAPGLPQYTGISEMTKPCPDIE-RSAFTVLSGLPLPFPKPII 77
DB 116 GPTRGPRGKGDGEGDPLPGV-----CRGSIYLSKAFSFGITTSPEERPII 166
OY 78 FTGVLVNAQRDLK EAMGVFACRVGNVYSSFDVELHCKVNIWILM 122
DB 167 FKKVLFNEGEHNPATGKRTCAFGIYFSYDITLANKHLAIGLV 211

RESULT 14
OY5JUD7 PRELIMINARY; PRT; 243 AA.

AC OY5JUD7.
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ADIPONECTIN.
GN APML.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopitheidae; Macaca.
RX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-ADIPOSE TISSUE;
RX MEDLINE=2123234; PubMed=11334417;
RA Hotta K., Funahashi T., Bodkin N.L., Ortmeier H.K., Arita Y.,
RA Hansen B.C., Matsuzawa Y.;
RT "Circulating concentrations of the adipocyte protein adiponectin are
RT decreased in parallel with reduced insulin sensitivity during the
RT progression to type 2 diabetes in rhesus monkeys."
RL Diabetes 50:1126-1133(2001).
DR EMBL: AF404407; AAK92202.1; -.
SQ SEQUENCE 243 AA; 26264 MM; 49A45DAF2B4613FD CRC64;

Query Match 5.6%; Score 140.5; DB 6; Length 243;
Best Local Similarity 28.8%; Pred. No. 0.00011;
Matches 40; Conservative 17; Mismatches 51; Indels 31; Gaps 3;

OY 17 VAGPPAHP-----RPPEEYGPAPGLPQYTGISE--MTKPCPD----- 55
DB 39 MAGTGHGHNHGVPRGDRGDRGEGKGRDGLGPRGDTGVTGAEGRGFRPFIQ 98
OY 56 -----IRSAFTVLSGLPLPFPKPIITFTGLVNAQRDLK EAMGVFACRVGNV 105

Db 99 RKGPBGCAVYNSAFSGVLETVYVPMPIRFTKIFYNQONHYDSTGKHCNIPGLY 158
QY 106 SSFDELHCKXNIMLRK 124
Db 159 FAYHITVYMKDVKSLFK 177

RESULT 15

062789 PRELIMINARY; PRT; 173 AA.
AC 062789;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE COLLAGEN VIII (FRAGMENT).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxId=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=AORTIC SMOOTH MUSCLE;
RA Reichenberg S., Plenz G., Robenek H.;
RL Submitted (MAR-1998) to the EMBL/genBank/DBJ databases.
DR EMBL; AF054891; AAC08999.1; -.
DR InterPro; IPR001073; Clq.
DR Pfam; PF00386; Clq; 1.
DR PRINTS; PRO0007; COMPLEMENTC1Q.
DR SMART; SM00110; Clq; 1.
DR PROSITE; PS01113; Clq; 1.
FT NON_TER 1 1
FT NON_TER 173 173
SQ SEQUENCE 173 AA; 18973 MW; 03489B6FBAB6CAD CRC64;

Query Match 5.5%; Score 137; DB 6; Length 173;
Best Local Similarity 30.3%; Pred. No. 0.00014;
Matches 37; Conservative 18; Mismatches 45; Indels 22; Gaps 5;

QY 20 PPAHPPEPEVG-----PGAPGLPQYTGISEMTKPCPDIERSAFTVKLSGKLPLP 72
Db 7 PPHGETLPDMGIGIDGAKPHAYG-----AKKGKNGSPAYEMPAFTAEITAPFPV 58
QY 73 FKPIIFGVLYNQRLKEANGVAFACRVPGNYSSFDELHCK-VNIVLMRKQIILANKE 131
Db 59 GAPVKEKLLYNGRQNYNPQGTGFTCEVPGVYFAYHV--HCKGNVWV---ALEKNN 112
QY 132 EI 133
Db 113 PV 114

Search completed: June 30, 2002, 11:54:29
Job time: 145 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2002, 11:52:24 ; Search time 13.41 Seconds

(Without alignments)
1325,300 Million cell updates/sec

Title: US-09-997-610-2

Sequence: 1 IVVIVPLITRAVIEHVEVACP.....GQFFPGTAVFTQCLYLHCM 459

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	330	13.2	196	1 HP20_TAMSI	Q06575 tamias sibi
2	221	8.9	215	1 HP25_TAMSI	Q06576 tamias sibi
3	171	6.9	743	1 CA18_MOUSE	Q00780 mus musculus
4	167.5	6.7	215	1 HP27_TAMSI	Q06577 tamias sibi
5	165.5	6.6	674	1 CA1A_CHICK	P08125 gallus galli
6	163.5	6.6	744	1 CA18_HUMAN	P27658 homo sapien
7	163.5	6.6	744	1 CA18_RABIT	P14282 oryctolagus
8	159.5	6.4	680	1 CA1A_MOUSE	Q05306 mus musculus
9	159	6.4	674	1 CA1A_MOUSE	P23206 bos taurus
10	156	6.3	680	1 CA1A_BOVIN	Q03692 homo sapien
11	152.5	6.1	247	1 APML_MOUSE	Q06094 mus musculus
12	148.5	6.0	244	1 APML_HUMAN	P15848 homo sapien
13	146.5	5.9	635	1 CA28_HUMAN	P25067 homo sapien
14	132.5	5.3	245	1 C10C_HUMAN	P02747 homo sapien
15	124.5	5.0	245	1 C10C_HUMAN	P075973 homo sapien
16	118.5	4.8	245	1 C10C_HUMAN	P075973 homo sapien
17	115.5	4.6	258	1 C10C_HUMAN	P18128 escherichia
18	111.5	4.5	1035	1 TAC2_MOUSE	O88992 mus musculus
19	111	4.5	255	1 GLIC_MOUSE	O91190 mus musculus
20	110.5	4.4	170	1 CA28_MOUSE	Q96874 mus musculus
21	110.5	4.4	251	1 C10B_HUMAN	P25318 mus musculus
22	110	4.4	3210	1 C10B_HUMAN	P02746 homo sapien
23	107.5	4.3	670	1 C10B_HUMAN	P49454 homo sapien
24	105.5	4.2	245	1 C10A_MOUSE	P54792 homo sapien
25	105	4.2	670	1 C10A_MOUSE	P98086 mus musculus
26	103	4.1	633	1 XRC1_HUMAN	I14640 homo sapien
27	101.5	4.1	551	1 XRC1_HUMAN	P18887 homo sapien
28	101	4.0	253	1 GFZF3_YEAST	P12944 saccharomyc
29	101	4.0	1026	1 C10B_MOUSE	P14106 mus musculus
30	100	4.0	245	1 TAC2_HUMAN	O95359 homo sapien
31	96.5	3.9	1005	1 C10A_HUMAN	P02745 homo sapien
32	96.5	3.9	1356	1 EAL2_HUMAN	O43491 homo sapien
33	96	3.8	246	1 C10C_MOUSE	O12767 homo sapien
					Q02105 mus musculus

34	95.5	3.8	763	1 FXM1_HUMAN	Q08050 h forkhead
35	95	3.8	1618	1 NEST_HUMAN	P48681 homo sapien
36	94.5	3.8	600	1 SP96_DICDI	P14328 dictyostell
37	94.5	3.8	628	1 V7OK_TYMC	P28478 turnip yell
38	94	3.8	695	1 DV1L_MOUSE	P51141 mus musculus
39	94	3.8	695	1 DV1L_MOUSE	O9wrb9 rattus norv
40	94	3.8	698	1 YB06_YEAST	P38283 saccharomyc
41	94	3.8	806	1 B1WA_EMENT	P17885 emericella
42	93.5	3.7	1237	1 B3A2_RABIT	P48746 oryctolagus
43	93	3.7	1162	1 LEPR_MOUSE	P48356 mus musculus
44	93	3.7	3149	1 TEGU_EBV	P03186 epstein-bar
45	92.5	3.7	253	1 C10B_RAT	P31721 rattus norv

ALIGNMENTS

```

RESULT 1
ID      HP20_TAMSI      STANDARD;      PRT;      196 AA.
AC      Q06575;
DT      01-NOV-1995 (Rel. 32, Created)
DT      01-NOV-1995 (Rel. 32, Last sequence update)
DT      01-MAR-2002 (Rel. 41, Last annotation update)
DE      Hibernation-associated plasma protein HP-20 precursor (Hibernator-
DE      specific blood complex, 20 kDa subunit).
OS      Tamias sibiricus (Siberian chipmunk) (Asian chipmunk).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sclerogathi; Scuriidae; Scuriinae;
OC      Tamias.
OX      NCBI_TaxID=64680;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Liver;
RX      MEDLINE=93180798; PubMed=8441393;
RA      Takamatsu N., Oba K., Kondo J., Kondo N., Shiba T.;
RT      "Hibernation-associated gene regulation of plasma proteins with a
RT      collagen-like domain in mammalian hibernators.";
RL      Mol. Cell. Biol. 13:1516-1521(1993).
RN      [2]
RP      SEQUENCE OF 24-58; 66-99; 104-129; 132-136 AND 150-184.
RC      TISSUE=Plasma;
RX      MEDLINE=92112696; PubMed=1730610;
RA      Kondo N., Kondo J.;
RT      "Identification of novel blood proteins specific for mammalian
RT      hibernation.";
RL      J. Biol. Chem. 267:473-478(1992).
CC      -I- FUNCTION: PLASMA PROTEINS HP-20, HP-25, HP-27 AND HP-55 FORM A
CC      140 kDa COMPLEX VIA DISULFIDE BONDS IN THE PLASMA AND ARE
CC      HIBERNATION SPECIFIC.
CC      -I- SUBCELLULAR LOCATION: Extracellular.
CC      -I- TISSUE SPECIFICITY: PLASMA; SYNTHESIZED IN THE LIVER.
CC      -I- DEVELOPMENTAL STAGE: THE PROTEIN COMPLEX DISAPPEARS FROM THE
CC      PLASMA AT ONSET OF HIBERNATION AND REAPPEARS AS HIBERNATION
CC      CEASES.
CC      -I- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC      -I- SIMILARITY: CONTAINS 1 C10 DOMAIN.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
DR      EMBL, D12974; BA02351.1; .
DR      InterPro, IPR001073; C1Q.
DR      InterPro, IPR000087; Collagen.
DR      Pfam, PF00386; C1Q; 1.
DR      PRINTS, PR00007; COMPLEMENTC1Q.
DR      SMART, SM01110; C1Q; 1.
DR      PROSITE, PS01113; C1Q; 1.

```

KW Signal; Collagen; Glycoprotein; Plasma; Multigene family.
FT SIGNAL 1 23
FT CHAIN 24 196 HIBERNATION-ASSOCIATED PLASMA PROTEIN
FT DOMAIN 25 63 HP-20.
FT DOMAIN 65 196 COLLAGEN-LIKE.
SQ SEQUENCE 196 AA; 21416 MW; CC3FA0A27854C6A2 CRC64;

Query Match 13.2%; Score 330; DB 1; Length 196;
Best Local Similarity 46.9%; Pred. No. 3.5e-18;
Matches 69; Conservative 21; Mismatches 45; Indels 12; Gaps 4;

QY 1 IYVIVPLINAVIEHVEVAGP---AHRRPEEVGP-----PGAGLPQYTGISEMTKC 51
DQ 7 LAIFVLMVNLNDQVSCSGPGPVGVPGRPGPGGPGAGRPDPGRKGSVVC 66
QY 52 PCPDIERSAFTYKLSGKLPPIFTGYLVYNAQRDLKEAMGVACRVNGYSSFDVE 111
DQ 67 PCR--ERSAFYKFSGRLLPPESEPVFTVLYNQRDLKESTGYFNCVPEGNTHFSFDVE 124
QY 112 LHCKVNIWLMRKQI-LANKEISIKQ 137
DQ 125 LYHCKVYIGLMKNHIOVMEKHQLSKNE 151

RESULT 2
HP25_TAMSI STANDARD; PRT; 215 AA.
ID HP25_TAMSI
AC Q06576;
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 41, Last annotation update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hibernation-associated plasma protein HP-25 precursor (Hibernator-specific blood complex, 25 kDa subunit).
OS Tamias sibiricus (Siberian chipmunk) (Asian chipmunk).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae;
OC Tamias.
NC NCBL_TaxID=64680;
NX NCBL_TaxID=64680;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=93180798; PubMed=8441393;
RA Takanatsu N., Ohba K., Kondo J., Kondo N., Shiba T.;
RT "Hibernation-associated gene regulation of plasma proteins with a collagen-like domain in mammalian hibernators.";
RL Mol. Cell. Biol. 13:1516-1521(1993).
RN [2]
RP SEQUENCE OF 29-62; 84-130; 172-183; 187-192 AND 201-215.
RC TISSUE=Plasma;
RX MEDLINE=92112696; PubMed=1730610;
RA Kondo N., Kondo J.;
RT "Identification of novel blood proteins specific for mammalian hibernation.";
RL J. Biol. Chem. 267:473-478(1992).
CC -|- FUNCTION: PLASMA PROTEINS HP-20, HP-25, HP-27 AND HP-55 FORM A 140 kDa COMPLEX VIA DISULFIDE BONDS IN THE PLASMA AND ARE HIBERNATION SPECIFIC.
CC -|- SUBCELLULAR LOCATION: Extracellular.
CC -|- TISSUE SPECIFICITY: PLASMA; SYNTHESIZED IN THE LIVER.
CC -|- DEVELOPMENTAL STAGE: THE PROTEIN COMPLEX DISAPPEARS FROM THE PLASMA AT ONSET OF HIBERNATION AND REAPPEARS AS HIBERNATION CEASES.
CC -|- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -|- SIMILARITY: CONTAINS 1 C1Q DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.
CC -----
DR EMBL: D12975; BAA02352.1; .
DR PIR: B48150; B48150.
DR InterPro: IPR001073; C1q.
DR InterPro: IPR000087; Collagen.
DR Pfam: PF00386; C1q; 1.
DR PRINTS: PRO0007; COMPLEMENTC1Q.
DR SMART: SM00110; C1Q; 1.
DR PROSITE: PS01113; C1Q; 1.
KW Signal; Collagen; Glycoprotein; Plasma; Multigene family.
FT SIGNAL 1 28
FT CHAIN 29 215 HIBERNATION-ASSOCIATED PLASMA PROTEIN
FT DOMAIN 40 81 HP-25.
FT DOMAIN 83 215 COLLAGEN-LIKE.
FT CARBOHYD 167 167 C1Q.
SQ SEQUENCE 215 AA; 22664 MW; AFE03206917EA530 CRC64;

Query Match 8.9%; Score 221; DB 1; Length 215;
Best Local Similarity 34.8%; Pred. No. 6.8e-10;
Matches 57; Conservative 20; Mismatches 69; Indels 18; Gaps 3;

QY 16 EVAGPPAHRRPP-----EVGPPGAPGLPQYTGISEMTKPCPDIERSAFTY 63
DQ 37 ECGPPGPPGPPGPIPGFPGAPGALGPPPGVPPIPGPQPGDVER--CSSRPKSAFAV 94
QY 64 KLSGKLPFPKPIFTGYLVYNAQRDLKEAMGVACRVNGYSSFDVELHCKVNIWLMR 123
DQ 95 KLSRRPPEPPQPIVFKEALYNQEGHFNMAIPEGFSCVLPGYNNGFDRLFGSSVYKRLMR 154
QY 124 KQILANKEI---SKQDSIQEVTVWLLKAFSTIREAHKSSN 163
DQ 155 DGIOVREKEAANDSYKHAMGSVIMALGDKVWLSEKLGKTES 198

RESULT 3
C1A8_MOUSE STANDARD; PRT; 743 AA.
ID C1A8_MOUSE
AC Q00780;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Collagen alpha 1(VIII) chain precursor.
DE COL8A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NX NCBL_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RX MEDLINE=92362626; PubMed=1499564;
RA Muragaki Y., Shiota C., Inoue M., Ooshima A., Olsen B.R., Nishimura Y.;
RT "Alpha 1(VIII)-collagen gene transcripts encode a short-chain collagen polypeptide and are expressed by various epithelial, endothelial and mesenchymal cells in newborn mouse tissues.";
RL Eur. J. Biochem. 207:895-902(1992).
CC -|- FUNCTION: MAJOR COMPONENT OF THE DESCERET'S MEMBRANE (BASEMENT MEMBRANE) OF CORNEAL ENDOTHELIAL CELLS.
CC -|- SUBUNIT: MAY FORM HOMOTRIMERS, OR HETEROTRIMERS IN ASSOCIATION WITH ALPHA 2(VIII) TYPE COLLAGENS.
CC -|- TISSUE SPECIFICITY: HIGH LEVELS IN CALVARIUM, EYE & SKIN OF NEWBORN MICE; ALSO IN VARIOUS EPITHELIAL, ENDOTHELIAL AND MESENCHYMAL CELLS.
CC -|- PTM: POLYLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -|- SIMILARITY: STRONG, TO ALPHA 2 TYPES VIII AND X COLLAGENS.
CC -|- SIMILARITY: CONTAINS 1 C1Q DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X66976; CAA47387.1; -
DR EMBL: X66977; CAA47387.1; JOINED.
DR PIR: S23779; S23779.
DR MGD: MGI:88463; Col8a1.
DR InterPro: IPR001073; C1q.
DR InterPro: IPR000087; Collagen.
DR Pfam: PF00386; C1q; 1.
DR Pfam: PF01391; Collagen; 7.
DR PRINTS: PRO0007; COMPLEMENTC1Q.
DR SMART: SM00110; C1Q; 1.
DR PROSITE: PS01113; C1Q; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Cell adhesion; Collagen; Signal.
FT SIGNAL 1 28
FT CHAIN 29 743 COLLAGEN ALPHA 1(VIII) CHAIN.
FT DOMAIN 29 117 NONHELICAL REGION (NC2).
FT DOMAIN 118 571 TRIPLE-HELICAL REGION (COL1).
FT DOMAIN 572 743 NONHELICAL REGION (NC1).
FT DOMAIN 608 743 C1Q.
SQ SEQUENCE 743 AA; 73595 MW; 350E29502653E0B CRC64;

Query Match 6.9%; Score 171; DB 1; Length 743;
Best Local Similarity 27.4%; Pred. No. 1.8e-05;
Matches 51; Conservative 26; Mismatches 65; Indels 44; Gaps 8;

QY 19 GPPHPRPPEVGPFGARGL---PQYGE-ISEM-----TKCP-----CPDIE 57
DB 554 GQGLPGRPGPPGPPGPPAVPTSPGSEYLPDMGLGIDGKTPATYAGKKKGPPAYE 613
QY 58 RSAFTVLSGLPLPFKPIIFTGVLNARDLKEAMGVFACRVPGNYSSFDVELHCK- 116
DB 614 MPATFAELTVFPFPGAVKFKDKLLYNGRQVNPQTGIFTCSEVPVYFAVHV--HCKG 670
QY 117 VNITLM---RKQILANKKEISKQSDIENVVWLKAFSIFREAHKSENNLPDNIKK 172
DB 671 GNVVVALFKNNEPMWYTYDEYK-----GFLDAQSGSAVLLLRPDQYFL 715
QY 173 KNFPE 178
DB 716 QNPFQ 721

RESULT 4
HP27_TAMSI STANDARD; PRT; 215 AA.
ID HP27_TAMSI
AC 006577;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hibernation-associated plasma protein HP-27 precursor (Hibernator-
DE specific blood complex, 27 kDa subunit).
OS Tamias sibiricus (Siberian chipmunk) (Asian chipmunk).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae;
OC Tamias.
OX NCBI_TaxID=6480;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver.
RX MEDLINE=93180798; PubMed=8441393;
RA Takamatsu N., Oba K., Kondo J., Kondo N., Shiba T.;
DE "Hibernation-associated gene regulation of plasma proteins with a
RT collagen-like domain in mammalian hibernators.";
RN Mol. Cell. Biol. 13:1516-1521(1993).
[2]

RP SEQUENCE OF 31-215.
RC TISSUE=Plasma;
RX MEDLINE=92112696; PubMed=1730610;
RA Kondo N., Kondo J.;
RT "Identification of novel blood proteins specific for mammalian
RT hibernation.";
RL J. Biol. Chem. 267:473-478(1992).
CC -I- FUNCTION: PLASMA PROTEINS HP-20, HP-25, HP-27 AND HP-55 FORM A
CC 140 kDa COMPLEX VIA DISULFIDE BONDS IN THE PLASMA AND ARE
CC HIBERNATION SPECIFIC.
CC -I- SUBCELLULAR LOCATION: Extracellular
CC -I- TISSUE SPECIFICITY: PLASMA; SYNTHESIZED IN THE LIVER.
CC -I- DEVELOPMENTAL STAGE: THE PROTEIN COMPLETELY DISAPPEARS FROM THE
CC PLASMA AT ONSET OF HIBERNATION AND REAPPEARS AS HIBERNATION
CC CEASES.
CC -I- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -I- SIMILARITY: CONTAINS 1 C1Q DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: D12976; BAA02353.1; -
DR InterPro: IPR001073; C1q.
DR InterPro: IPR000087; Collagen.
DR Pfam: PF00386; C1q; 1.
DR PRINTS: PRO0007; COMPLEMENTC1Q.
DR SMART: SM00110; C1Q; 1.
DR PROSITE: PS01113; C1Q; 1.
KW Signal; collagen; Glycoprotein; Plasma; Multigene family.
FT SIGNAL 1 30
FT CHAIN 31 215
FT DOMAIN 43 81 HIBERNATION-ASSOCIATED PLASMA PROTEIN
FT DOMAIN 83 215 HP-27.
FT CARBOHYD 155 155 COLLAGEN-LIKE.
FT COMFLICT 41 155 C1Q.
FT COMFLICT 51 51 N-LINKED (GLCNAC...) (POTENTIAL).
FT CONFLICT 51 51 V -> A (IN REF. 2).
FT CONFLICT 51 51 R -> Q (IN REF. 2).
SQ SEQUENCE 215 AA; 22797 MW; D2357086A08BFA9E CRC64;

Query Match 6.7%; Score 167.5; DB 1; Length 215;
Best Local Similarity 31.5%; Pred. No. 7.5e-06;
Matches 51; Conservative 14; Mismatches 68; Indels 29; Gaps 4;

QY 6 VLTAVIEHV-----EVAGP-----PAHPRPEEVGPFGARGLPQY 41
DB 18 ILALSVLHVVCSETQGNPDESCNVPGQPGPMGRPGPGRGPPGNMGFGRLGPPGP 77
QY 42 TGEISEMTKCPDIERSAFTYKLSGLPLPFKPIIFTGVLNARDLKEAMGVFACRV 101
DB 78 PG---MT-VNCHSKSGTSAFAVKANELPPAPSPVIFKEALHDAGHFDLAVETCPV 132
QY 102 GNVYSSFDVELHCKVNIWLMRKQILANKKEISKQSDIENV 143
DB 133 GLYQGFHIEAVQRAVKYSLNRNGTVQVNEREAQDGYEHIS 174

RESULT 5
CALA_CHICK STANDARD; PRT; 674 AA.
ID CALA_CHICK
AC P08125;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Collagen alpha 1(X) chain precursor.
GN COL10A1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_Taxid=9031;
RN [1]
RP SEQUENCE OF 48-674 FROM N.A., AND SEQUENCE OF 103-117 AND 453-466.
RX MEDLINE=86168227; Pubmed=3082876;
RA Ninomiya Y., Gordon M., van der Nest M., Schmid T., Linsemayer T.,
Olsen B.R.;
RT "The developmentally regulated type X collagen gene contains a long
RT open reading frame without introns.";
RL J. Biol. Chem. 261:5041-5050(1986).
RN [2]
RP SEQUENCE OF 1-75 FROM N.A.
RX MEDLINE=89054019; Pubmed=2461368;
RA Luvalle P., Ninomiya Y., Rosenblum N.D., Olsen B.R.;
RT "The type X collagen gene. Intron sequences split the 5'-untranslated
RT region and separate the coding regions for the non-collagenous amino-
RT terminal and triple-helical domains.";
RL J. Biol. Chem. 263:18378-18385(1988).
RN [3]
RP REVISIONS TO C-TERMINUS.
RX MEDLINE=89380199; Pubmed=2476437;
RA Yamaguchi N., Benya P.D., van der Rest M., Ninomiya Y.;
RT "The cloning and sequencing of alpha 1(VIII) collagen cDNAs
RT demonstrate that type VIII collagen is a short chain collagen and
RT contains triple-helical and carboxyl-terminal non-triple-helical
RT domains similar to those of type X collagen.";
RL J. Biol. Chem. 264:16022-16029(1989).
CC -1- FUNCTION: TYPE X COLLAGEN IS A PRODUCT OF HYPERTROPHIC
CC CHONDROCYTES AND HAS BEEN LOCALIZED TO PRESUMPTIVE
CC MINERALIZATION ZONES OF HYALINE CARTILAGE.
CC -1- SUBUNIT: HOMOTRIMER.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- SIMILARITY: STRONG, TO ALPHA 1 AND 2 TYPE VIII COLLAGENS.
CC -1- SIMILARITY: CONTAINS 1 C10 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M13456; AAA48736.1; ALT_SEQ.
CC EMBL: J04194; AAA48634.1; -.
CC PIR: A31896; A31896.
CC InterPro: IPR001073; C1q.
CC InterPro: IPR000087; Collagen.
CC Pfam: PF00386; C1q; 1.
CC Pfam: PF01391; Collagen; 6.
CC PRINTS: PR00007; COMPLEMENTC1Q.
CC SMART: SM00110; C1Q; 1.
CC PROSITE: PS00113; C1Q; 1.
CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
CC Cartilage; Collagen; Signal.
FT SIGNL 1 18
FT CHAIN 19 674 COLLAGEN ALPHA 1(X) CHAIN.
FT DOMAIN 19 52 NONHELICAL REGION (NC2).
FT DOMAIN 53 512 TRIPLE-HELICAL REGION.
FT DOMAIN 513 674 NONHELICAL REGION (NC1).
FT DOMAIN 539 674 C1Q.
FT MOD_RES 453 453 HYDROXYLATION.
FT MOD_RES 456 456 HYDROXYLATION.
SO SEQUENCE 674 AA; 66434 MW; EBA48B1EF174B145 CRC64;

Query Match 6.6%; Score 165.5; DB 1; Length 674;
Best Local Similarity 36.4%; Pred. No. 4.1e-05;
Matches 44; Conservative 11; Mismatches 41; Indels 25; Gaps 5;

1 QY 18 AGPAPHPPEVGPAGALPO--YTGELISE-----MTKCPDIERSAF 61

Db 494 SCPEGLPPGPPGPGGTTIEGVKGESEKSGSMFKAGANQALGMPV-----SAF 548
QY 62 TVKLSGKLPLPFKPIIFGVLYNAORDLKEAMGVACFACVGNYSFDELHCK-VNITW 120
Db 549 TVLSATAYGATVPKFKDKILYNRQHDPTGTGIFCRIGLYPFSTYH---HAKGTNW 605
QY 121 L 121
Db 606 V 606
RESULT 6
CAL8_HUMAN STANDARD; PRT; 744 AA.
ID CAL8_HUMAN
AC P27658;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 1(VIII) chain precursor (Endothelial collagen).
GN COL8A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9131001; Pubmed=2029894;
RA Nureaki Y., Mattei M.-G., Yamaguchi N., Olsen B.R., Ninomiya Y.;
RT "The complete primary structure of the human alpha 1 (VIII) chain and
RT assignment of its gene (COL8A1) to chromosome 3.";
RL Eur. J. Biochem. 197:615-622(1991).
CC -1- FUNCTION: MAJOR COMPONENT OF THE DESCMET'S MEMBRANE (BASEMENT
CC MEMBRANE) OF CORNEAL ENDOTHELIAL CELLS.
CC -1- SUBUNIT: MAY FORM HOMOTRIMERS, OR HETEROTRIMERS IN ASSOCIATION
CC WITH ALPHA 2(VIII) TYPE COLLAGENS.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- MISCELLANEOUS: FOUR CONSECUTIVE GLY-PRO-PRO TRIPLETS ARE PRESENT
CC AT THE C-TERMINUS OF THE TRIPLE-HELICAL REGION. THESE MAY PROVIDE
CC THE HIGH THERMAL STABILITY OF THIS REGION.
CC -1- SIMILARITY: STRONG, TO ALPHA 2 TYPES VIII AND X COLLAGENS.
CC -1- SIMILARITY: CONTAINS 1 C10 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X57527; CAA40748.1; -.
CC PIR: S15435; S15435.
CC MIM: 120251;
CC InterPro: IPR001073; C1q.
CC InterPro: IPR000087; Collagen.
CC Pfam: PF00386; C1q; 1.
CC Pfam: PF01391; Collagen; 7.
CC PRINTS: PR00007; COMPLEMENTC1Q.
CC SMART: SM00110; C1Q; 1.
CC PROSITE: PS00113; C1Q; 1.
CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
CC Glycoprotein; Cell adhesion; Collagen; Signal.
FT SIGNL 1 28
FT CHAIN 29 744 COLLAGEN ALPHA 1(VIII) CHAIN.
FT DOMAIN 29 117 NONHELICAL REGION (NC2).
FT DOMAIN 118 571 TRIPLE-HELICAL REGION (COL1).
FT DOMAIN 572 744 NONHELICAL REGION (NC1).
FT DOMAIN 609 744 C1Q.
SO SEQUENCE 744 AA; 73443 MW; 7852C22582B377A3 CRC64;

Query Match 6.6%; Score 163.5; DB 1; Length 744;
 Best Local Similarity 30.0%; Pred. No. 6.6e-05;
 Matches 42; Conservative 17; Mismatches 48; Indels 33; Gaps 5;

19 GPPAHPREPEVGPAGL-----PQYGEISEMKKCCP 54
 Db 554 GQPLGPPGPPGPPGPPAVMPPTPQGEYLPMGLGIDGVKPPHAG--AKKKNCGP 611
 QY 55 DIERSAFVKSGLPLPEFKPIFTGVLYNAORDLKEMGVACRPVGNYSFDELHH 114
 Db 612 AYEMPAFTALTPFPVGVGPKFNKLLYNGRONYNPQTGTFCEVGYFFAYHV---H 668
 QY 115 CK-VNIWLMRKQILANKKEI 133
 Db 669 CKGNVWV---ALFKNNEPV 685

RESULT 7
 CALA_RABIT STANDARD; PRT; 744 AA.
 AC P14282;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Collagen alpha 1(VIII) chain precursor (Endothelial collagen).
 GN COL8A1.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=69380199; PubMed=2476437;
 RA Yamaguchi N., Benya P.D., van der Rest M., Ninomiya Y.;
 RT "The cloning and sequencing of alpha 1(VIII) collagen cDNAs
 demonstrates that type VIII collagen is a short chain collagen and
 contains triple-helical and carboxyl-terminal non-triple-helical
 domains similar to those of type X collagen.";
 RL J. Biol. Chem. 264:16022-16029(1989).
 CC -1- FUNCTION: MAJOR COMPONENT OF THE DESCMET'S MEMBRANE (BASEMENT
 CC MEMBRANE) OF CORNEAL ENDOTHELIAL CELLS.
 CC -1- SUBUNIT: MAY FORM HOMOTRIMERS, OR HETEROTRIMERS IN ASSOCIATION
 CC WITH ALPHA 2(VIII) TYPE COLLAGENS.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- MISCELLANEOUS: FOUR CONSECUTIVE GLY-PRO TRIPEPTS ARE PRESENT
 CC AT THE C-TERMINUS OF THE TRIPLE-HELICAL REGION. THESE MAY PROVIDE
 CC HIGH THERMAL STABILITY OF THIS REGION OF THIS REGION.
 CC -1- SIMILARITY: STRONG. NO ALPHA 2 TYPES VIII AND X COLLAGENS.
 CC -1- SIMILARITY: CONTAINS 1 C10 DOMAIN.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: J05042; AAA31204.1; -.
 DR PIR: A34246; A34246.
 DR InterPro: IPR001073; C1q.
 DR InterPro: IPR000087; Collagen.
 DR Pfam: PF00386; C1q; 1.
 DR Pfam: PF01391; Collagen; 7.
 DR PRINTS: PR00007; COMPLEMENTC1Q.
 DR SMART: SM00110; C1q; 1.
 DR PROSITE: PS01113; C1q; 1.
 KM Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Cell adhesion; Collagen; Signal.
 FT SIGNAL 1 20
 FT CHAIN 21 744 COLLAGEN ALPHA 1(VIII) CHAIN.
 FT DOMAIN 29 117 NONHELICAL REGION (NC2).

FT DOMAIN 118 571 TRIPLE-HELICAL REGION.
 FT DOMAIN 572 744 NONHELICAL REGION (NC1).
 FT DOMAIN 609 744 C1Q.
 SQ SEQUENCE 744 AA; 73358 MW; 2A8CE1E8274E99 CMC64;

Query Match 6.6%; Score 163.5; DB 1; Length 744;
 Best Local Similarity 32.6%; Pred. No. 6.6e-05;
 Matches 45; Conservative 16; Mismatches 48; Indels 29; Gaps 7;

19 GPPAHPREPEVGPAGL-----PQYGE-ISEM-----TKCP-----CPDI 56
 Db 554 GQPLGPPGPPGPPGPPAVMPPTPQGEYLPMGLGIDGVKPPHAYAAKKNCGPAY 613
 QY 57 ERSAFVKSGLPLPEFKPIFTGVLYNAORDLKEMGVACRPVGNYSFDELHCK 116
 Db 614 EMPAFTALTPFPVGVGPKFIDRLLYNGRONYNPQTGTFCEVGYFFAYHV---HCK 670
 QY 117 -VNIWLMRKQILANKKEI 133
 Db 671 GGNVWV---ALFKNNEPV 685

RESULT 8
 CALA_MOUSE STANDARD; PRT; 680 AA.
 AC Q05306;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Collagen alpha 1(X) chain precursor.
 GN COL10A1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/C;
 RX MEDLINE=93143676; PubMed=8424763;
 RA Elima K., Berola I., Rosati R., Metsaranta M., Garofalo S., Perala M.,
 RA de Crombrughe B., Vuorio E.;
 RT "The mouse collagen X gene: complete nucleotide sequence, exon
 RT structure and expression pattern.";
 RL Biochem. J. 289:247-253(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SV; TISSUE=Liver;
 RX MEDLINE=93238750; PubMed=8477738;
 RA Kong R.Y.C., Kwai K.M., Lau E.T., Thomas J.T., Boot-Handford R.P.,
 RA Grant M.E., Cheah K.S.E.;
 RT "Intron-exon structure, alternative use of promoter and expression of
 RT the mouse collagen X gene, Col10a-1.";
 RL Eur. J. Biochem. 213:99-111(1993).
 RN [3]
 RP SEQUENCE OF 51-680 FROM N.A.
 RC STRAIN=DBA/2J;
 RX MEDLINE=92267014; PubMed=1587271;
 RA Ape S.S., Seldin M.F., Hayashi M., Olsen B.R.;
 RT "Cloning of the human and mouse type X collagen genes and mapping of
 RT the mouse type X collagen gene to chromosome 10.";
 RL Eur. J. Biochem. 206:217-224(1992).
 RN [4]
 RP SEQUENCE OF 385-627 FROM N.A.
 RC STRAIN=C57BL;
 RX MEDLINE=92182017; PubMed=1543751;
 RA Elima K., Metsaranta M., Kallio J., Peraelae M., Berola I.,
 RA Garofalo S., de Crombrughe B., Vuorio E.;
 RT "Specific hybridization probes for mouse alpha 2(Ix) and alpha 1(X)
 RT collagen mRNAs.";
 RL Biochim. Biophys. Acta 1130:78-80(1992).
 CC -1- FUNCTION: TYPE X COLLAGEN IS A PRODUCT OF HYPERTROPHIC
 CC CHONDROCYTES AND HAS BEEN LOCALIZED TO PRESUMPTIVE

```
CC MINERALIZATION ZONES OF HYALINE CARTILAGE.
CC -1- SUBUNIT: HOMOTRIMER.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- SIMILARITY: STRONG, TO ALPHA 1 AND 2 TYPE VIII COLLAGENS.
CC -1- SIMILARITY: CONTAINS 1 C10 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X67348; CAA47763.1; -
DR EMBL; X65121; CAA46237.1; -
DR EMBL; X63013; CAA44741.1; -
DR EMBL; X21610; CAA79736.1; -
DR PIR; S28807; S28807.
DR PIR; S31216; S31216.
DR PIR; S22215; S22215.
DR MGD; MGI:88445; Col10a1.
DR InterPro; IPR001073; C1q.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF00386; C1q; 1.
DR PRINTS; PR00007; COLLAGEN.6.
DR SMART; SM00110; C1Q; 1.
DR PROSITE; PS01113; C1Q; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Cartilage; Collagen; Signal.
FT SIGNAL 1 18
FT CHAIN 19 680 COLLAGEN ALPHA 1(X) CHAIN.
FT DOMAIN 19 56 NONHELICAL REGION (NC2).
FT DOMAIN 57 519 TRIPLE-HELICAL REGION.
FT DOMAIN 520 680 NONHELICAL REGION (NC1).
FT DOMAIN 545 680 C1Q.
FT CONFLICT 248 248 P -> L (IN REF. 3).
FT CONFLICT 286 286 A -> S (IN REF. 2).
FT CONFLICT 306 306 L -> F (IN REF. 3).
FT CONFLICT 417 417 T -> S (IN REF. 3).
FT CONFLICT 451 451 R -> K (IN REF. 4).
FT CONFLICT 500 500 H -> L (IN REF. 3).
FT CONFLICT 567 567 A -> C (IN REF. 3).
FT CONFLICT 569 569 I -> H (IN REF. 3).
FT CONFLICT 571 572 FD -> IY (IN REF. 3).
FT CONFLICT 635 635 Q -> T (IN REF. 3).
SQ SEQUENCE 680 AA; 66775 MW; FE984CA99AE708E2 CRC64;

Query Match 6.4%; Score 159.5; DB 1; Length 680;
Best Local Similarity 29.6%; Pred. No. 0.00012;
Matches 55; Conservative 18; Mismatches 68; Indels 45; Gaps 8;
```

```
CA1A_BOVIN
ID CA1A_BOVIN STANDARD. PRT; 674 AA.
AC P23206;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE collagen alpha 1(X) chain precursor.
GN COL10A1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=cartilage;
RX MEDLINE=91113131; PubMed=1703407;
RA Thomas J.T., Kwan A.P.L., Grant M.E., Boot-Handford R.P.;
RT "Isolation of cDNAs encoding the complete sequence of bovine type X
RT collagen. Evidence for the condensed nature of mammalian type X
RT collagen genes."
RL Biochem. J. 273:141-148(1991).
CC -1- FUNCTION: TYPE X COLLAGEN IS A PRODUCT OF HYPERTHROPHIC
CC CHONDROCYTES AND HAS BEEN LOCALIZED TO PRESUMPTIVE
CC MINERALIZATION ZONES OF HYALINE CARTILAGE.
CC -1- SUBUNIT: HOMOTRIMER.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- SIMILARITY: STRONG, TO ALPHA 1 AND 2 TYPE VIII COLLAGENS.
CC -1- SIMILARITY: CONTAINS 1 C10 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X53556; CAA37624.1; -
DR PIR; S13301; S13301.
DR InterPro; IPR001073; C1q.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF00386; C1q; 1.
DR PRINTS; PR00007; COLLAGEN.7.
DR SMART; SM00110; C1Q; 1.
DR PROSITE; PS01113; C1Q; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Cartilage; Collagen; Signal; Glycoprotein.
FT SIGNAL 1 18
FT CHAIN 19 674 COLLAGEN ALPHA 1(X) CHAIN.
FT DOMAIN 19 56 NONHELICAL REGION (NC2).
FT DOMAIN 57 519 TRIPLE-HELICAL REGION.
FT DOMAIN 520 674 NONHELICAL REGION (NC1).
FT DOMAIN 539 674 C1Q.
FT DISULFID 194 197 BY SIMILARITY. (BY SIMILARITY).
FT MOD_RES 460 460 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 463 463 HYDROXYLATION (BY SIMILARITY).
FT CARBOHYD 611 611 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 674 AA; 65546 MW; CD4CA73A03E004CA CRC64;

Query Match 6.4%; Score 159; DB 1; Length 674;
Best Local Similarity 31.8%; Pred. No. 0.00013;
Matches 41; Conservative 16; Mismatches 56; Indels 16; Gaps 3;
```

RESULT 9

DB 556 YPAIGTPFDKLVNKQOHYDPRGTFCTKIPGTYFESHVKTGHAWGLYKNGTPV 615
 OY 127 LANKKEISK 135
 DB 616 MYTDEXIK 624
 RESULT 10
 CALA_HUMAN STANDARD; PRT: 680 AA.
 ID CALA_HUMAN
 AC 003692;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Collagen alpha 1(X) chain precursor.
 GN COL10A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92109659; PubMed-1764025;
 RA Thomas J.T., Cresswell C.J., Rash B., Nicolai H., Jones T.,
 RA Solomon E., Grant M.E., Boot-Handford R.P.;
 RT "The human collagen X gene. Complete primary translated sequence and
 RT chromosomal localization.";
 RL Biochem. J. 280:617-623(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-93012005; PubMed-1397333;
 RA Reichenberger E., Beler F., Luvaille P., Olsen B.R., von der Mark K.,
 RA Bertling W.M.;
 RT "Genomic organization and full-length cDNA sequence of human collagen
 RT X.";
 RL FEBS Lett. 311:305-310(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Beler F., Lamm M.B., von der Mark K.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Williams S.;
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 52-680 FROM N.A.
 RX MEDLINE-92267014; PubMed-1587271;
 RA Apte S.S., Seidlin M.F., Hayashi M., Olsen B.R.;
 RT "Cloning of the human and mouse type X collagen genes and mapping of
 RT the mouse type X collagen gene to chromosome 10.";
 RL Eur. J. Biochem. 206:217-224(1992).
 RN [6]
 RP SEQUENCE OF 561-666 FROM N.A.
 RX MEDLINE-91243838; PubMed-2037056;
 RA Apte S., Mattei M.-G., Olsen B.R.;
 RT "Cloning of human alpha 1(X) collagen DNA and localization of the
 RT COL10A1 gene to the q21-q22 region of human chromosome 6.";
 RL FEBS Lett. 282:393-396(1991).
 RN [7]
 RP SEQUENCE OF 547-655 FROM N.A.
 RX MEDLINE-92077285; PubMed-1743401;
 RA Reichenberger E., Aigner T., von der Mark K., Stoeb H., Bertling W.;
 RT "In situ hybridization studies on the expression of type X collagen
 RT in fetal human cartilage.";
 RL Dev. Biol. 148:562-572(1991).
 RN [8]
 RP REVIEW ON VARIANTS.
 RX MEDLINE-97255959; PubMed-9101290;
 RA Kuitvanleem H., Tromp G., Prockop D.J.;
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibrill-
 RT associated collagen (type IX), and network-forming collagen (type X)
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
 RL Hum. Mutat. 9:300-315(1997).

RN [9]
 RP VARIANTS SMCD ASP-598 AND PRO-614.
 RX MEDLINE-94136476; PubMed-8304336;
 RA Wallis G.A., Rash B., Sweetman W.A., Thomas J.T., Super M., Evans G.,
 RA Grant M.E., Boot-Handford R.P.;
 RT "Amino acid substitutions of conserved residues in the
 RT carboxyl-terminal domain of the alpha 1(X) chain of type X collagen
 RT occur in two unrelated families with metaphyseal chondrodysplasia
 RT type Schmid.";
 RL Am. J. Hum. Genet. 54:169-178(1994).
 RN [10]
 RP VARIANT SMCD ARG-591.
 RX MEDLINE-94272470; PubMed-8004099;
 RA McIntosh I., Abbott M.H., Warman M.L., Olsen B.R., Francomano C.A.;
 RT "Additional mutations of type X collagen confirm COL10A1 as the
 RT Schmid metaphyseal chondrodysplasia locus.";
 RL Hum. Mol. Genet. 3:303-307(1994).
 RN [11]
 RP VARIANT SMCD VAL-618.
 RX MEDLINE-95181449; PubMed-7876225;
 RA Chan D., Cole W.G., Rogers J.G., Bateman J.F.;
 RT "Type X collagen multimer assembly in vitro is prevented by a Gly618
 RT to Val mutation in the alpha 1(X) NCI domain resulting in Schmid
 RT metaphyseal chondrodysplasia.";
 RL J. Biol. Chem. 270:4558-4562(1995).
 RN [12]
 RP VARIANTS SMCD ARG-545; GLU-595; HIS-597; LYS-617; ARG-644 AND GLY-648.
 RX MEDLINE-95331767; PubMed-7607655;
 RA Bonaventure J., Chaminade F., Maroteaux P.;
 RT "Mutations in three subdomains of the carboxy-terminal region of
 RT collagen type X account for most of the Schmid metaphyseal
 RT dysplasias.";
 RL Hum. Genet. 96:58-64(1995).
 RN [13]
 RP VARIANT SMCD PRO-600.
 RX MEDLINE-96375754; PubMed-8782043;
 RA Wallis G.A., Rash B., Sykes B., Bonaventure J., Maroteaux P.,
 RA Zabel B., Wynne-Davies R., Grant M.E., Boot-Handford R.P.;
 RT "Mutations within the gene encoding the alpha 1 (X) chain of type X
 RT collagen (COL10A1) cause metaphyseal chondrodysplasia type Schmid but
 RT not several other forms of metaphyseal chondrodysplasia.";
 RL J. Med. Genet. 33:450-457(1996).
 RN [14]
 RP VARIANTS SMCD GLU-18 AND ARG-18.
 RX MEDLINE-97220591; PubMed-9067753;
 RA Ikegawa S., Nakamura K., Nagano A., Haga N., Nakamura Y.;
 RT "Mutations in the N-terminal globular domain of the type X collagen
 RT gene (COL10A1) in patients with Schmid metaphyseal
 RT chondrodysplasia.";
 RL Hum. Mutat. 9:131-135(1997).
 RN [15]
 RP VARIANTS SMD GLU-595.
 RX MEDLINE-99057503; PubMed-9837818;
 RA Ikegawa S., Nishimura G., Nagai T., Hasegawa T., Ohashi H.,
 RA Nakamura Y.;
 RT "Mutation of the type X collagen gene 'COL10A1' causes
 RT spondylometaphyseal dysplasia.";
 RL Am. J. Hum. Genet. 63:1659-1662(1998).
 RN [16]
 RP VARIANT SMCD CYS-597.
 RX MEDLINE-99069781; PubMed-9852679;
 RA Sawai H., Ida A., Nakata Y., Koyama K.;
 RT "Novel missense mutation resulting in the substitution of tyrosine by
 RT cysteine at codon 597 of the type X collagen gene associated with
 RT Schmid metaphyseal chondrodysplasia.";
 RL J. Hum. Genet. 43:259-261(1998).
 CC -I- FUNCTION: TYPE X COLLAGEN IS A PRODUCT OF HYPERTROPHIC
 CC CHONDROCYTES AND HAS BEEN LOCALIZED TO PRESUMPTIVE
 CC MINERALIZATION ZONES OF HYALINE CARTILAGE.
 CC -I- SUBUNIT: HOMOTRIMER.
 CC -I- PTM: PROLINES ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -I- UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -I- DISEASE: DEFECTS IN COL10A1 ARE THE CAUSE OF SCHMID TYPE

	CC	METAPHASEAL CHONDRODYSPLASIA(SMCD), A DOMINANTLY INHERITED DISORDER OF THE OSSEOUS SKELETON. THE CARDINAL FEATURES OF THE PHENOTYPE ARE MILD SHORT STATURE, COXA VARA AND A WADDLING GAIT.
	CC	RADIOGRAPH USUALLY SHOWS SCLEROSIS OF THE RIBS, FLARING OF THE METAPHASES, AND A WIDE IRREGULAR GROWTH PLATE, ESPECIALLY OF THE KNEES.
	CC	-1- DISEASE DEFECTS IN COL10A1 ARE THE CAUSE OF SPONDILOMETAPHYSEAL DYSPLASIA(SMD). SMD COMPRISES A HETEROGENEOUS GROUP OF HERITABLE SKELETAL DYSPHASIAS CHARACTERIZED BY MODIFICATIONS OF THE VERTICAL BODIES OF THE SPINE AND METAPHASES OF THE TUBULAR BONES.
	CC	-1- SIMILARITY: STRONG, TO ALPHA 1 AND 2 TYPE VIII COLLAGENS.
	CC	-1- SIMILARITY: CONTAINS 1 C1Q DOMAIN.
	CC	-----
	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to licenses@isb.sib.ch).
	CC	-----
	DR	EMBL; X60382; CAA42933.1; -.
	DR	EMBL; X65120; CAA46236.1; -.
	DR	EMBL; X98568; CAA67178.1; -.
	DR	EMBL; AL121963; CAB87590.1; -.
	DR	EMBL; S68531; AAC60615.1; -.
	DR	EMBL; X58879; CAA41686.1; -.
	DR	EMBL; M74050; AAA61221.1; -.
	DR	EMBL; X72579; CAA51170.1; -.
	DR	EMBL; X72580; CAA51170.1; JOINED.
	DR	PIR; S15626; S15826.
	DR	PIR; S30086; S30086.
	DR	PIR; A43901; A43901.
	DR	PIR; S18249; S18249.
	DR	PIR; S21856; S21856.
	DR	PIR; S26396; S26396.
	DR	MIM; 120110; -.
	DR	MIM; 156500; -.
	DR	MIM; 184250; -.
	DR	IncePro; IPR001073; Clqg.
	DR	IncePro; IPR000087; Collagen.
	DR	Pfam; PF00386; Clq; 1.
	DR	Pfam; PFG01391; Collagen; 7.
	DR	PRINTS; PR00007; COMPLENKTClO.
	DR	SMART; SMART0110; ClO; 1.
	KW	PROSITE; PS01113; ClO; 1.
	KW	Extradellular matrix; Connective tissue; Repeat; Hydroxylation; Cariliagen; Collagen; Signal; Disease mutation; Polymorphism.
	FT	SIGNAL 1 18
	FT	CHAIN 19 680
	FT	DOMAIN 19 56 NONHELICAL REGION (NC2).
	FT	DOMAIN 57 519 TRIPLE-HELICAL REGION.
	FT	DOMAIN 520 680 NONHELICAL REGION (NC1).
	FT	VARIANT 545 680
	FT	18 18
	FT	G->E (IN SMCD).
	FT	VARIANT 18 18
	FT	G->R (IN SMCD).
	FT	G->R (IN SMCD).
	FT	VARIANT 545 545
	FT	G->R.
	FT	VARIANT 591 591
	FT	/FTId=VAR..001840.
	FT	C->R (IN SMCD).
	FT	/FTId=VAR..001841.
	FT	VARIANT 595 595
	FT	G->E (IN SMCD AND SMD).
Oy		Query Match 6.3%; Score 156; DB 1; Length 680;
		Best Local Similarity 33.9%; Pred. No. 0.00022;
		Matches 39; Conservative 16; Mismatches 46; Indels 14; Gaps 4;
Ddb	18 AGPAHPRPPEEGPGADLPQ---YTGEISEMTKPCPDIER-----SAFYTKSG 67 : : : : :	
	501 SGEGGLPGRPPGPPGGCAVMEGETIKAGQRSLSGTPLVSANOGVTGMPSAFVIISK 560 :	
Oy	68 KLDPFKPFIIFTGVLYNAORDLKEMAGVFACRVNGNTSSDFDELHHCK-VNIWL 121 :	

DB	APML	MOUSE	STANDARD	PRT	247 AA
RESULT 11					
APML					
ID	APML				
AC	060994	062400	09DC68		
DT	01-NOV-1997	(Rel. 35, Created)			
DT	01-NOV-1997	(Rel. 35, Last sequence update)			
DE	16-OCT-2001	(Rel. 40, Last annotation update)			
DR	Adiponectin precursor (30 kDa adipocyte complement-related protein)				
DE	(ACRP30) (adipocyte specific protein Adipoq).				
GN	APML OR ACRP30 OR ADIPOQ.				
OS	Mus musculus (Mouse).				
OC	Eumetazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RA	SEQUENCE FROM N.A.				
RC	TISSUE=Adipocyte;				
RC	MEDLINE=96070757; PubMed=7592907;				
RC	Scherer P.E., Williams S., Fogliano M., Baldini G., Lodish H.F.;				
RT	"A novel serum protein similar to C1q, produced exclusively in				
RT	adipocytes."				
RL	J. Biol. Chem. 270:26746-26749(1995).				
RP	[2]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Fibroblast;				
RC	MEDLINE=96209999; PubMed=8631877;				
RC	Hu E., Liang P., Spiegelman B.M.;				
RT	"Adipoq is a novel adipose-specific gene dysregulated in obesity."				
RT	J. Biol. Chem. 271:10697-10703(1996).				
RL	[3]				
RP	SEQUENCE FROM N.A.				
RC	Das K., Lin Y., Widen E., Zhang Y., Scherer P.E.;				
RC	PubMed=11162643;				
RT	"Chromosomal localization, expression pattern, and promoter analysis				
RT	of the mouse gene encoding adipocyte-specific secretory protein				
RT	Acip30."				
RL	Biochem. Biophys. Res. Commun. 280:1120-1129(2001).				
RL	[4]				
RP	SEQUENCE FROM N.A.				
RC	SPRAIN=C57BL/6J; TISSUE=Heart;				
RC	MEDLINE=21085660; PubMed=11217851;				
RC	Kawai T., Shinaigawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,				
RA	Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,				
RA	Aizawa K., Iizawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,				
RA	Saito T., Okazaki Y., Gotohori T., Bono H., Kusakawa T., Saito R.,				
RA	Kodora K., Matsuda H.A., Ashburner M., Batlora S., Casavant T.,				
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochia H.,				
RA	Knehl P., Lewis S., Matsuo Y., Nakato I., Pesole G., Quackenbush J.,				
RA	Schirni L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,				
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,				
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,				
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,				
RA	Guntonich S., Hill D., Hofmann M., Hume D.A., Kamya M., Lee N.H.,				
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Nomberts P.,				
RA	Nordone P., Ring B., Ringstad M., Rodriguez I., Sakamoto N.,				
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,				
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Winking L.,				
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,				
RA	Hayashizaki Y.;				
RT	"Functional annotation of a full-length mouse cDNA collection."				
RT	Nature 409:685-690(2001).				
RL	FUNCTION: IMPORTANT NEGATIVE REGULATOR IN HEMATOPOIESIS AND IMMUNE				
CC	SYSTEMS; MAY BE INVOLVED IN ENDING INFLAMMATORY RESPONSES THROUGH				
CC	ITS INHIBITORY FUNCTIONS. INHIBITS ENDOTHELIAL NF-KAPPA SIGNALING				
CC	THROUGH A CAMP-DEPENDENT PATHWAY. INHIBITS TNF-ALPHA-INDUCED				
CC	EXPRESSION OF ENDOTHELIAL ADHESION MOLECULES (BY SIMILARITY).				
CC	-1- SUBUNIT: HOMODIGOMER.				
CC	-1- SUBCELLULAR LOCATION: SECRETED IN PLASMA.				
CC	-1- TISSUE SPECIFICITY: SYNTHESIZED EXCLUSIVELY BY ADIPOCYTES AND				
CC	-1- TISSUE SPECIFICITY: SYNTHESIZED EXCLUSIVELY BY ADIPOCYTES AND				

```
CC SECRETED INFO SERUM.
CC -1- INDUCTION: DURING HORMONE-INDUCED ADIPOSE DIFFERENTIATION AND
CC ACTIVATED BY INSULIN.
CC -1- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 C1Q DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U37222; AAA80543.1; -
DR EMBL: U49915; AB06706.1; -
DR EMBL: AF304466; AAK13417.1; -
DR EMBL: AK00138; BAB22597.1; -
DR MGD: MGI:106675; Acip30.
DR InterPro: IPR001073; C1q.
DR Pfam: PF00386; C1q; 1.
DR Pfam: PF01391; Collagen; 1.
DR PRINTS: PR00007; COMPLEMENTC1Q.
DR SMART: SM0110; C1Q; 1.
DR PROSITE: PS01113; C1Q; 1.
KW Collagen; Signal; Repeat; Hydroxylation; Plasma; Polymorphism.
FT SIGNAL 1 17
FT CHAIN 18 247
FT DOMAIN 45 110
FT DISULFID 39 39
FT MOD_RES 47 47
FT MOD_RES 50 50
FT MOD_RES 56 56
FT MOD_RES 65 65
FT MOD_RES 79 79
FT MOD_RES 98 98
FT MOD_RES 107 107
FT VARIANT 113 113
FT CONFLICT 50 50
FT CONFLICT 74 74
FT CONFLICT 117 117
FT CONFLICT 148 148
FT CONFLICT 243 243
SQ SEQUENCE 247 AA; 26841 MW; 137B687D873988C4 CRC64;

Query Match 6.1%; Score 152.5; DB 1; Length 247;
Best Local Similarity 29.7%; Pred. No. 0.00012;
Matches 44; Conservative 23; Mismatches 64; Indels 17; Gaps 4;

QY 31 GPPGAPGLPQYVNGEISEMKKPCQPDIERSAFTYKLSGKILPFRKPIIFPGVLYNAGRD LK 90
DB 93 GRRGPPGPTGRRGGEPEEAAV-----MYRSATFSGLTFRVTPVNPFRFKIFENQGNHYD 147
QY 91 EAMGAFACHPGPNYSSPDEVLH--HCKYNIWLMRKQILANKKEISQSQIOEVTWVLK 148
DB 148 GSTGKPYCNICPLYYFSYIITYMKDKVYSLFKRKDAKAVLFTDQYQEKNVDAQSSSVLLH 207
QY 149 A-----FSFIREAEHKSSENLPDNNV 169
DB 208 LEVGDQVWLQVYGGDDH---NGLYADNV 232

RESULT 12
APML_HUMAN STANDARD: PRT; 244 AA.
AC Q15848;
DC 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DI 01-MAR-2002 (Rel. 41, Last annotation update)
DE Adiponectin precursor (30 kDa adipocyte complement-related protein)
```

```
DE (ACRP30) (adipose most abundant gene transcript 1) (apm-1) (gelatin-
DE binding protein).
GN APML OR ACRP30 OR GBP28.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Adipose tissue;
RX MEDLINE=96224171; PubMed=8619847;
RA Maeda K., Okubo K., Shimomura I., Funahashi T., Matsuzawa Y.,
RA Matsubara K.;
RT "cDNA cloning and expression of a novel adipose specific collagen-like
RT factor, apml (Adipose Most abundant Gene transcript 1).";
RL Biochem. Biophys. Res. Commun. 221:286-289(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99196984; PubMed=10095105;
RA Saito K., Tobe T., Minoshima S., Asakawa S., Sumiya J., Yoda M.,
RA Nakano Y., Shimizu N., Tomita M.;
RT "Organization of the gene for gelatin-binding protein (GBP28).";
RL Gene 229:67-73(1999).
RN [3]
RP SEQUENCE FROM N.A.
RA Schaeffler A., Orso E., Paltzsch K.D., Buechler C., Drobnik W.,
RA Fuernst A., Schaefer J., Schmitz G.;
RT "The human apm-1, an adipocyte-specific gene linked to the family of
RT TNF's and to genes expressed in activated T cells, is mapped to
RT chromosome 1q21.3-q23, a susceptibility locus identified for familial
RL combined hyperlipidemia (FCH).";
RL Biochem. Biophys. Res. Commun. 260:416-425(1999).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=20417747; PubMed=10961870;
RA Yokota T., Oritani K., Takahashi T., Ishikawa J., Matsuyama A.,
RA Uuchi N., Kihara S., Funahashi T., Tenner A.J., Tomiyama Y.,
RA Matsuzawa Y.;
RT "Adiponectin, a new member of the family of soluble defense collagens,
RT negatively regulates the growth of myelomonocytic progenitors and the
RL functions of macrophages.";
RL Blood 96:1723-1732(2000).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=20440368; PubMed=10982546;
RA Uuchi N., Kihara S., Arita Y., Okamoto Y., Maeda K., Kuriyama H.,
RA Hotta K., Nishida M., Takahashi M., Muraguchi M., Ohmoto Y.,
RA Nakamura T., Yamashita S., Funahashi T., Matsuzawa Y.;
RT "Adiponectin, an adipocyte-derived plasma protein, inhibits
RT endothelial NF-kappaB signaling through a camp-dependent pathway.";
RL Circulation 102:1296-1301(2000).
CC -1- FUNCTION: IMPORTANT NEGATIVE REGULATOR IN HEMATOPOIESIS AND IMMUNE
CC SYSTEMS; MAY BE INVOLVED IN ENDING INFLAMMATORY RESPONSES THROUGH
CC ITS INHIBITORY FUNCTIONS. INHIBITS ENDOTHELIAL NF-KAPPAB SIGNALING
CC THROUGH A CAMP-DEPENDENT PATHWAY. INHIBITS TNF-ALPHA-INDUCED
CC EXPRESSION OF ENDOTHELIAL ADHESION MOLECULES.
CC -1- SUBUNIT: HOMODIGOMER (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: SECRETED IN PLASMA.
CC -1- TISSUE SPECIFICITY: EXPRESSED ONLY IN THE ADIPOSE TISSUE.
CC -1- SIMILARITY: CONTAINS 1 C1Q DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D45371; BAA08227.1; -
DR EMBL: AB012165; BAA86716.1; -
DR EMBL: AB012164; BAA86716.1; JOINED.
```

DR EMBL: AJ131460; CAB52413.1; -.
 DR MIM: AJ131461; CAB52413.1; JOINED.
 DR MIM: 605441; -.
 DR InterPro: IPR001073; C1q.
 DR InterPro: IPR000087; Collagen.
 DR Pfam: PF00386; C1q; 1.
 DR Pfam: PF01391; Collagen; 1.
 DR PRINTS: PRO0007; COMPLEMENTC1Q.
 DR SMART: SM00110; C1q; 1.
 DR PROSITE: PS01113; C1q; 1.
 DR Collagen; Signal; Repeat; Hydroxylation; Plasma.
 KW SIGNAL 1 14 POTENTIAL.
 FT CHAIN 15 244 ADIPONECTIN.
 FT DOMAIN 42 107 COLLAGEN-LIKE.
 FT DOMAIN 108 244 C1Q.
 FT DISULFID 36 36 INTERCHAIN (BY SIMILARITY).
 FT MOD_RES 44 44 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 47 47 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 53 53 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 62 62 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 71 71 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 76 76 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 86 86 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 95 95 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 104 104 HYDROXYLATION (BY SIMILARITY).
 SQ SEQUENCE 244 AA; 26414 MW; 64D8C6C1204B1018 CRC64;

Query Match 6.0%; Score 148.5; DB 1; Length 244;
 Best Local Similarity 30.7%; Pred. No. 0.00024;
 Matches 43; Conservative 16; Mismatches 48; Indels 33; Gaps 4;

QY 17 VAGPPAP-----RPEVGPAGAPGLPYTGELISEMTPKCP-----CPDIE 57
 DB 40 MAGIPGPHNGAPGADGDPGEGKEGDPGLGPCKDIGE-TGVPAEGRGEPGIG 98
 QY 58 -----RSAPFYKLSGKLPFPKPIFTGVLYNAQRDLKAMGVFACRPVGN 104
 DB 99 GRKGEGEGAYVRSAPSVGLETYTIRMPRTFKIRTYNQNNHYDSTGKFCNIPGLX 158
 QY 105 YSSEFVELHCKKYNIMLRK 124
 DB 159 YFAHITVYMKDVKSLFKK 178

RESULT 13
 CA28_HUMAN STANDARD; PRT; 635 AA.
 ID CA28_HUMAN
 AC P25067;
 DT 01-MAY-1992 (Rel. 22, Last Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Collagen alpha 2(VIII) chain (Endothelial collagen) (Fragment).
 GN COL8A2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91210292; PubMed=2019595;
 RA Muragaki Y., Jacenko O., Apte S., Mattel M.-G., Ninomiya Y.,
 RA Olsen B.R.;
 RT "The alpha 2(VIII) collagen gene. A novel member of the short chain
 RT collagen family located on the human chromosome 1.";
 RL J. Biol. Chem. 266:7721-7727(1991).
 CC -!- FUNCTION: MAJOR COMPONENT OF THE DESCENDING'S MEMBRANE (BASEMENT
 CC MEMBRANE) OF CORNEAL ENDOTHELIAL CELLS.
 CC -!- SUBUNIT: MAY FORM HOMOTRIMERS, OR HETEROTRIMERS IN ASSOCIATION
 CC WITH ALPHA 1(VIII) TYPE COLLAGENS.
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -!- SIMILARITY: STRONG, TO ALPHA 1 TYPES VIII AND X COLLAGENS.

CC -!- SIMILARITY: CONTAINS 1 C1Q DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 DR EMBL: M60832; AAA62822.1; -.
 DR MIM: 120252; -.
 DR InterPro: IPR001073; C1q.
 DR InterPro: IPR000087; Collagen.
 DR Pfam: PF00386; C1q; 1.
 DR Pfam: PF01391; Collagen; 7.
 DR SMART: SM00110; C1q; 1.
 DR PROSITE: PS01113; C1q; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Cell adhesion; Collagen.
 FT NON_TER 1 1
 FT DOMAIN <1 11 NONHELICAL REGION (NC2).
 FT DOMAIN 12 468 TRIPLE-HELICAL REGION.
 FT DOMAIN 469 635 NONHELICAL REGION (NC1).
 FT DOMAIN 500 635 C1Q.
 SQ SEQUENCE 635 AA; 60527 MW; 22A261164754F771 CRC64;

Query Match 5.9%; Score 146.5; DB 1; Length 635;
 Best Local Similarity 31.0%; Pred. No. 0.001;
 Matches 40; Conservative 9; Mismatches 51; Indels 29; Gaps 4;

QY 19 GPAPHPPEVGPAGAPGLPYTGELISEM-----TKCPDIERSS----- 59
 DB 442 GPPEVPSPGPIITPPELPAPGAPGAFDETGIAGLHPNGGEGAVLYGKGGKPGFGLGEI 501
 QY 60 -----AFYTKLSGKLPFPKPIFTGVLYNAQRDLKAMGVFACRPVGN 113
 DB 502 SAHATPAFPAVLTLSPLPASGMFPAKFDRTLYNGHSGYNPAITGIFCPGVGYFAHYH--- 558
 QY 114 HCK-VNIML 121
 DB 559 HVKGTNVWV 567

RESULT 14
 C1QC_HUMAN STANDARD; PRT; 245 AA.
 ID C1QC_HUMAN
 AC P02747;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Complement C1q subcomponent, C chain precursor.
 GN C1QG OR C1QC.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Monocytes;
 RX MEDLINE=91174759; PubMed=1706597;
 RA Sellier G.C., Blake D.J., Reid K.B.M.;
 RT "Characterization and organization of the genes encoding the A-, B-
 RT and C-chains of human complement subcomponent C1q. The complete
 RT derived amino acid sequence of human C1q.";
 RL Biochem. J. 274:481-490(1991).
 RN [2]
 RP SEQUENCE OF 29-122.
 RX MEDLINE=80020137; PubMed=486087;
 RA Reid K.B.M.;
 RT "Complete amino acid sequences of the three collagen-like regions
 RT present in subcomponent C1q of the first component of human

Search completed: June 30, 2002, 11:54:49
Job time: 145 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2002, 11:51:29 ; Search time 20.93 Seconds

(without alignments)
2107.263 Million cell updates/sec

Title: US-09-997-610-2

Perfect score: 2494

Sequence: 1 IVVIVPLITRAVIEHVEVAGP.....GQFFPFGTAVFTQCLYLHGM 459

Scoring table: BLOSUM62

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	330	13.2	196	2	A48150 hibernation-relate
2	221	8.9	215	2	A48150 hibernation-relate
3	171	6.9	743	1	S23779 collagen alpha 1(V
4	167.5	6.7	215	2	C48150 hibernation-relate
5	166.5	6.6	744	1	S23298 collagen alpha 1(V
6	165.5	6.6	674	2	S23297 collagen alpha 1(V
7	163.5	6.6	744	1	A34246 collagen alpha 1(V
8	163.5	6.6	744	2	S15435 collagen alpha 1(V
9	159.5	6.4	680	2	S13216 collagen alpha 1(V
10	159	6.4	674	2	S13301 collagen alpha 1(V
11	156	6.3	680	1	CGH01D collagen alpha 1(X
12	148.5	6.0	244	2	JC4708 collagen alpha 1(X
13	146.5	5.9	635	2	A57131 gelatin-binding 28
14	132.5	5.3	245	1	C1H0UC collagen alpha 2(V
15	129.5	5.2	1553	2	C1H0UC complement subcomp
16	121.5	4.9	219	2	T03301 rab3 effector prot
17	117.5	4.7	245	2	S12089 hypothetical prote
18	111	4.5	898	2	S69634 hypothetical prote
19	110.5	4.4	170	2	B57131 collagen alpha 2(V
20	110.5	4.4	253	1	C1H0UC complement subcomp
21	110	4.4	813	2	S70795 vasa protein precu
22	110	4.4	1017	2	PC4035 cell-cycle-depende
23	106.5	4.3	1008	2	T9470 phosphatidic acid-
24	105.5	4.2	245	2	S19018 complement subcomp
25	103	4.1	405	4	A61181 homeotic protein H
26	103	4.1	633	2	A36353 DNA repair protein
27	101.5	4.1	551	2	S53377 G2f3 protein - yea
28	101	4.0	253	2	T49560 complement C1q B C
29	100.5	4.0	1280	2	T00365 hypothetical prote

30	100	4.0	245	1	C1H0UA complement subcomp
31	99.5	4.0	1495	2	T31434 densin-180 - rat
32	99	4.0	343	2	H64491 hypothetical prote
33	98.5	3.9	992	2	T08772 hypothetical prote
34	97.5	3.9	740	2	S17925 polynucleotide ade
35	97	3.9	207	1	O08CU2 hypothetical 21.9K
36	97	3.9	355	2	T29932 hypothetical prote
37	96.5	3.9	2187	2	T30826 nascent polypeptid
38	96	3.8	246	2	S29328 complement subcomp
39	96	3.8	620	2	T50150 yeast nrd1-like pr
40	96	3.8	920	2	T43263 cell division prot
41	95.5	3.8	895	2	S74225 leptin receptor, i
42	95	3.8	1618	2	S21424 nestin - human
43	94.5	3.8	481	2	G75253 hypothetical prote
44	94.5	3.8	600	2	S07638 spore coat protein
45	94.5	3.8	628	2	S19150 hypothetical prote

ALIGNMENTS

```
RESULT 1
A48150
hibernation-related protein HP-20 precursor - siberian chipmunk
C:Species: Eutamias sibiricus (Siberian chipmunk)
C:Date: 16-Feb-1994 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: A48150; C41752
R:Takamatsu, N.; Ohba, K.; Kondo, J.; Kondo, N.; Shiba, T.
Mol. Cell. Biol. 13, 1516-1521, 1993
A:Title: Hibernation-associated gene regulation of plasma proteins with a collagen-11
A:Reference number: A48150; MUID:93180798
A:Accession: A48150
A:Molecule type: mRNA
A:Residues: 1-196 <TAK>
A:Cross-references: GB:D12974; NID:9287467; PIDN:BA02351.1; PID:9287468
A:Note: the source is designated as Tamias asiaticus in Genbank entry TMSHP20A, relea
R:Kondo, N.; Kondo, J.
J. Biol. Chem. 267, 473-478, 1992
A:Title: Identification of novel blood proteins specific for mammalian hibernation.
A:Reference number: A41752; MUID:92112696
A:Accession: C41752
A:Status: preliminary
A:Molecule type: protein
A:Residues: 24-58;66-99;104-129;132-136;137;151-184 <KON>
C:Superfamily: complement subcomponent C1q chain A; complement C1q carboxyl-terminal
C:Keywords: glycoprotein; hibernation; plasma
F:1-24/Domain: signal sequence #status predicted <SIG>
F:24-196/Product: hibernation-related protein HP-20 #status experimental <MAT>
F:73-196/Domain: complement C1q carboxyl-terminal homology <C1Q>
F:154/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 13.2%; Score 330; DB 2; Length 196;
Best Local Similarity 46.9%; Pred. No. 1.6e+18;
Matches 69; Conservatively 21; Mismatches 45; Indels 12; Gaps 4;

OY 1 IVVIVPLITRAVIEHVEVAGP---AHRPPEEVGP-----CGAPLPQYTGISEMTKC 51
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7 LAIFLVNVLNDVSCSGPPGPGVGVGPGPGPGPGPGPGPGPGPGPGPGPGPGPG 66
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 52 PCEDIRSAFTYKLSGLDLPFRPIIFGTGLVYNAORDLEAMGVACRVPGNYSSFDVE 111
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 67 PCR--BRSAFTYKFSGRLLPPESEPVVETLVLYTORDLKESGVFCVPGNHFSDVE 124
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 112 LHHCKYNIWLAKQOT-LANKEELSKQO 137
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 125 LYHCKYIGLKNHIOVMEKHOLSKNE 151
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 2
B48150
hibernation-related protein HP-25 precursor - siberian chipmunk
```


A:Gene: GDB:COL8A1
A:Cross-references: GDB:128104; OMIM:120251
A:Map position: 3q11.1-3q13.2
C:Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-744/Product: collagen alpha 1(VIII) chain #status predicted <MAT>
F:21-117/Region: amino-terminal nonhelical
F:118-571/Region: interrupted helical
F:572-744/Region: carboxyl-terminal nonhelical
F:617-743/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match 6.68; Score 163.5; DB 2; Length 744;
Best Local Similarity 30.08; Pred. No. 9.4e-05;
Matches 42; Conservative 17; Mismatches 48; Indels 33; Gaps 5;

OY 19 GPPAHPRPPEEYGPAGCLP-----PQYGEISEMTKCPDIER-----
Db 554 GQPGLPFGPPGPPGPPAHPRPPEEYGPAGCLPDMGLGIDGVKPPHATG--AKKGNKGP 611
OY 55 DIERSAFYVLSGKLPDPKPIFTGVLYNAQRDLKEMGVFACRVPGNYSFDELHH 114
Db 612 AYEMPAFTALFTAPPPGPGVKNKLLYNQRYNPFQTFCEVPGYVFAVHV---H 668
OY 115 CK-VNITLMRKQILANKKEI 133
Db 669 CKGNVWV---ALFRNNEPV 685

RESULT 9
S31216
collagen alpha 1(X) chain precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 13-Aug-1999
C:Accession: S31216; S28807; S22215; S30127; I48299; S26397; S31830
R:Kong, R.Y.C.; Kwan, K.M.; Lau, E.T.; Thomas, J.T.; Boot-Handford, R.P.; Grant, M.E.; C
Eur. J. Biochem. 213, 99-111, 1993
A:Title: Intron-exon structure, alternative use of promoter and expression of the mouse
A:Reference number: S31216; MUID:93238750
A:Accession: S31216
A:Molecule type: DNA
A:Residues: 1-680 <KON>
A:Cross-references: EMBL:X21610; NID:g49793; PIDN:CAA79736.1; PID:g49794
R:Elima, K.; Eerola, I.; Rosati, R.; Metsaeranta, M.; Garofalo, S.; Petreia, M.; de Crc
Biochem. J. 289, 247-253, 1993
A:Title: The mouse collagen X gene: complete nucleotide sequence, exon structure and exp
A:Reference number: S28807; MUID:93143676
A:Accession: S28807
A:Molecule type: DNA
A:Residues: 1-285, 'A', 287-680 <ELI>
A:Cross-references: EMBL:X67348; NID:g50480; PIDN:CAA47763.1; PID:g50481
R:Elima, K.; Metsaeranta, M.; Kallio, J.; Petreia, M.; Eerola, I.; Garofalo, S.; de Crc
Biochim. Biophys. Acta 1130, 78-80, 1992
A:Title: Specific hybridization probes for mouse alpha-2(IX) and alpha-1(X) collagen mRN
A:Reference number: S22215; MUID:92182017
A:Accession: S22215
A:Molecule type: mRNA
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 385-450, 'K', 452-627 <ELIA>
A:Cross-references: EMBL:X63013; NID:g49795; PIDN:CAA44741.1; PID:g49796
R:Appte, S.S.; Olsen, B.R.
Matrix 13, 165-179, 1993
A:Title: Characterization of the mouse type X collagen gene.
A:Reference number: S30127; MUID:93261348
A:Accession: S30127
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-12, 'F', 14-26, 'S', 28-247, 'V', 249-285, 'A', 287-305, 'F', 307-416, 'S', 418-499, 'L
R:Appte, S.S.; Seldin, M.F.; Hayashi, M.; Olsen, B.R.
Eur. J. Biochem. 206, 217-224, 1992
A:Title: Cloning of the human and mouse type X collagen genes and mapping of the mouse t
A:Reference number: I48299; MUID:92267014
A:Accession: I48299

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 52-247, 'L', 249-285, 'A', 287-305, 'F', 307-416, 'S', 418-499, 'L', 501-566, 'C', 56
A:Cross-references: EMBL:X65121; NID:g50482; PIDN:CAA46237.1; PID:g667031
R:Summers, T.A.; Irwin, M.H.; Mayne, R.; Bellan, G.
J. Biol. Chem. 263, 581-587, 1988
A:Title: Monoclonal antibodies to type X collagen. Biosynthetic studies using an anti
A:Reference number: S26397; MUID:88087150
A:Accession: S26397
A:Molecule type: protein
A:Residues: 'SDGYFSQ', 24-26, 'KQ' <SUK>
A:Genetics:
A:Gene: COL10A-1
A:Map position: 10
A:Introns: 51/3
C:Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology
C:Keywords: collod coll; extracellular matrix; glycoprotein; homotrimer
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-680/Product: collagen alpha 1(X) chain #status predicted <MAT>
F:553-679/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match 6.48; Score 159.5; DB 2; Length 680;
Best Local Similarity 29.64; Pred. No. 0.00017;
Matches 55; Conservative 18; Mismatches 68; Indels 45; Gaps 8;

OY 19 GPPAHPRP-----PEEYGPAGCLPQY-----TGEISEMTKCPDIER----- 58
Db 490 GPPGPPGPGHSGEPGLPGPGPGQAVMPDGFTRKAGQPRSLGSANHGVTGM 549
OY 59 --SAFTYVLSGKLPDPKPIFTGVLYNAQRDLKEMGVFACRVGNYSSDELH--H 114
Db 550 PVSAFTYVLSKAYPAVGAIPDEILYNQOHVDPKSGIFGKIGIYFVHNVKGTG 609
OY 115 CKNVITLMRKQILANKKEISK-----QOSIOEV-----VWLKASFIFEAHK--- 159
Db 610 VVWGLXKKNQTPKMTYDYSKGYLDQASGSAIMETENDQVWLQLP-----NMSNGLY 663
OY 160 SSENELH 165
Db 664 SSEYVH 669

RESULT 10
S13301
collagen alpha 1(X) chain precursor - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 21-Nov-1993 #sequence_revision 23-Feb-1996 #text_change 13-Aug-1999
C:Accession: S13301
R:Thomas, J.T.; Kwan, A.P.L.; Grant, M.E.; Boot-Handford, R.P.
Biochem. J. 273, 141-148, 1991
A:Title: Isolation of cDNAs encoding the complete sequence of bovine type X collagen.
A:Reference number: S13301; MUID:91113131
A:Accession: S13301
A:Molecule type: mRNA
A:Residues: 1-674 <THO>
A:Cross-references: EMBL:X53556; NID:g263; PIDN:CAA37624.1; PID:g264
C:Genetics:
A:Gene: COL10A1
C:Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology
C:Keywords: collod coll; extracellular matrix; glycoprotein; homotrimer
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-674/Product: collagen alpha 1(X) chain #status predicted <MAT>
F:547-673/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match 6.48; Score 159; DB 2; Length 674;
Best Local Similarity 31.88; Pred. No. 0.00019;
Matches 41; Conservative 16; Mismatches 56; Indels 16; Gaps 3;

OY 18 AAPPAHPRPPEEYGPAGCLPQY-----YTGEISEMTKCPDIERSAFYVLSGK 68
Db 501 AGEPLPGPPGPPGPGQVALPDEYKAGORPVSANQCVTGMV-----SAFTYVLSKA 555

Oy 69 LPLPKPIIFGVLYINAGRDLEKMGVFCACRVCPGVYSSFPVELCH--HCKNIIMLMRKQI 126
 Db 556 YPAICGTPIPLFDKILYLNKQKHDPDRIGTICKIPGLIYYSYHIVHKGTAMVGLIKNCIPV 615
 Oy 127 LANKEEISK 135
 Db 616 MYTYDEVYIK 624
 RESULT 11
 CGHVID
 collagen alpha 1(X) chain precursor - human
 N:Alternate names: procollagen alpha 1(X) chain
 C:Species: Homo sapiens (man)
 C:Date: 22-Nov-1993 #sequence, revision 03-Nov-1995 #text change 22-Jun-1999
 C:Accession: S26396; S30086; S15826; S18249; A43901; I51870; S21856
 R:Reichenberger, E.; Belter, F.; Luvallie, P.; Disen, B.R.; von der Mark, K.; Bertling, W.
 FEBS Lett. 311, 305-310, 1992
 A:Title: Genomic organization and full-length cDNA sequence of human collagen X.
 A:Reference number: S26396; MUID:93012005
 A:Accession: S26396
 A:Molecule type: DNA
 A:Residues: 1-680 <REL>
 A:Cross-references: EMBL:X68952; EMBL:X72578; EMBL:X72579; EMBL:X72580; GB:S47714; GB:S47715; S30086; S15826; S18249; A43901; I51870; S21856
 R:Apte, S.S.
 submitted to the EMBL Data Library, March 1992
 A:Reference number: S30085
 A:Accession: S30086
 A:Molecule type: DNA
 A:Residues: 71PFYGVWCWVCL, 52-680 <AP2>
 A:Cross-references: EMBL:X65120; NID:923129
 A:Note: the initial difference is probably due to translation of an intronic sequence
 R:Apte, S.; Mattei, M.G.; Olsen, B.R.
 FEBS Lett. 282, 393-396, 1991
 A:Title: Cloning of human alpha-1(X) collagen DNA and localization of the CD10A1 gene to chromosome 12p11.2
 A:Reference number: S15826; MUID:91243838
 A:Accession: S15826
 A:Molecule type: DNA
 A:Residues: 561-647, 'G', 649-666 <AP2>
 A:Cross-references: EMBL:X58879; NID:930013; PIDN:CAA1686.1; PID:930014
 R:Thomas, J.T.; Cresswell, C.J.; Rash, B.; Nicolai, H.; Jones, T.; Solomon, E.; Grant, M.
 Biochem. J. 280, 617-623, 1991
 A:Title: The human collagen X gene: complete primary translated sequence and chromosomal localization
 A:Reference number: S18249; MUID:92109559
 A:Accession: S18249
 A:Molecule type: DNA
 A:Residues: 1-26, 'T', 28-680 <THD>
 A:Cross-references: EMBL:X60382; NID:930094; PIDN:CAA4293.1; PID:930095
 A:Note: the sequence from Fig. 3 is inconsistent with that from Fig. 2 in having 532-Ala
 R:Reichenberger, E.; Aigner, T.; von der Mark, K.; Stoss, H.; Bertling, W.
 Dev. Biol. 148, 562-572, 1991
 A:Title: In situ hybridization studies on the expression of type X collagen in fetal human cartilage
 A:Reference number: A43901; MUID:92077285
 A:Accession: A43901
 A:Molecule type: mRNA
 A:Residues: 547-656 <REL>
 A:Cross-references: GB:R4050; GB:D57494; NID:9339884; PIDN:AAA61221.1; PID:9553796
 A:Note: sequence extracted from NCBI backbone (NCBI:69012, NCBI:69014)
 R:Wallis, G.A.; Rash, B.; Sweetman, W.A.; Thomas, J.T.; Super, M.; Evans, G.; Grant, M.H.
 Am. J. Hum. Genet. 54, 169-178, 1994
 A:Title: Amino acid substitutions of conserved residues in the carboxyl-terminal domain of human collagen X
 A:Reference number: I51870; MUID:94136476
 A:Accession: I51870
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 520-597, 'D', 599-680 <MAL>
 A:Cross-references: GB:568531; NID:9545180; PIDN:AAC60615.1; PID:9545181
 A:Note: mutant sequence from patient with metaphyseal chondrodysplasia type Schmid
 C:Note: a second mutant sequence with 614-Pro is also described
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit are conserved and subsequently O-glycosylated.


```

Db 53 MLKCVVRDMAKPA---ACKTPRNAES-----QPHQPLNIF---RCVCPVRKPSSEGGP 101
OY 141 EYTWVLLKAFSFIREAHEKSENUL-----HPDNV-----IKKNPFSEGEKFLA-----E 186
Db 102 ERDMRHLHQOFESYKEGVRIKIGEARRYOGEHKKDDAPTGCICHTKTFADGGHLCSTYCRTK 161
OY 187 ICT-CNEELNVNPQDNGENISWTCQ--RSSQOSI-KSLAMPRRKRWFCGTGP-----GS 236
Db 162 FCARCGGRVSLRSNMEDKVMVMVNCNLCRKQOEILTCKSGA-----WFFGSGPQOPSDGT 215
OY 237 LCCVQ-----PRDLVPCV-----PVNSAVAS--EGASP-----KPMQLPS 269
Db 216 LSDTATGAGSEVPREKKARLOERSRSQTFPLSTAAYSQDTATPGAPLHRNKGAEPSQAL 275
OY 270 GVEPVGAKKSRIEVMPEPIRFQKIYGNPMPMPROKFAVGVSS-WRTSARYVOKGNVWEP 328
Db 276 GPEOKQASRSRS---EPPREKKA---PGLSEQNGKGQKSERKRVPKSVVQPG----- 323
OY 329 PHRVPSGAPSSRAVRRSPSSRLQKGRSTDSLQHVPEK 366
Db 324 -----EGIADERERKRRRETRETRLEKGRSQDYSDR-PEK 355

```

Search completed: June 30, 2002, 11:53:51
 Job time: 142 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2002, 11:50:59 ; Search time 16.33 seconds
(without alignments)
686,549 Million cell updates/sec

Title: US-09-997-610-2
Perfect score: 2494
Sequence: 1 IIVIPVLITAVIEHVEVAGP.....GQFFPGTAVFTQCLYLHCM 459

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2.6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2.6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2.6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2.6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2.6/ptodata/2/1aa/PCITUS.COMB.pep.*
6: /cgn2.6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	221	8.9	215	4	US-09-140-804-5
2	174	7.0	236	4	US-09-140-804-6
3	167	6.7	185	2	US-08-463-911-3
4	152.5	6.1	247	2	US-08-463-911-2
5	148.5	6.0	244	2	US-08-463-911-7
6	148.5	6.0	244	4	US-09-140-804-3
7	144.5	5.8	247	4	US-09-140-804-8
8	144.5	5.8	247	4	US-09-118-408-3
9	124	5.0	243	4	US-09-188-930-285
10	123.5	5.0	294	4	US-09-188-930-294
11	121.5	4.9	243	4	US-09-140-804-2
12	116	4.7	130	4	US-09-485-316A-12
13	114	4.6	130	4	US-09-485-316A-13
14	110	4.4	2482	1	US-08-328-254-6
15	110	4.4	3248	1	US-08-353-700-1
16	110	4.4	3248	5	PCT-US95-16216-1
17	105	4.2	130	4	US-09-485-316A-11
18	105	4.2	132	2	US-08-463-911-5
19	100	4.0	245	4	US-09-140-804-4
20	98	3.9	281	4	US-09-118-408-44
21	97	3.9	1261	3	US-09-080-855-2
22	96	3.8	246	2	US-08-463-911-4
23	95.5	3.8	763	2	US-08-742-753-4
24	95	3.8	1618	1	US-07-853-913-4
25	94	3.8	1754	1	US-07-745-206A-13
26	94	3.8	1754	2	US-08-311-363-13
27	93	3.7	894	2	US-08-599-455B-2

28	93	3.7	894	4	US-09-069-781B-2	Sequence 2, Appl
29	93	3.7	894	4	US-08-618-957A-12	Sequence 12, Appl
30	93	3.7	896	2	US-08-640-389A-12	Sequence 12, Appl
31	93	3.7	1162	2	US-08-599-455B-43	Sequence 43, Appl
32	93	3.7	1162	4	US-09-069-781B-43	Sequence 43, Appl
33	92.5	3.7	1847	6	5256642-10	Patent No. 5256642
34	92.5	3.7	1847	6	5472939-10	Patent No. 5472939
35	92.5	3.7	2039	6	5256642-2	Patent No. 5256642
36	92.5	3.7	2039	6	5472939-2	Patent No. 5472939
37	91	3.6	562	2	US-08-973-675-2	Sequence 2, Appl
38	91	3.6	1251	5	PCR-US95-02251-3	Sequence 3, Appl
39	91	3.6	1252	1	US-08-199-780-3	Sequence 3, Appl
40	91	3.6	1252	2	US-08-316-650-3	Sequence 3, Appl
41	91	3.6	1253	3	US-08-479-722B-4	Sequence 4, Appl
42	90.5	3.6	778	6	5198347-4	Patent No. 5198347
43	90.5	3.6	895	4	US-08-827-962-19	Sequence 19, Appl
44	90.5	3.6	895	4	US-08-827-962-21	Sequence 21, Appl
45	90.5	3.6	1162	4	US-08-827-962-15	Sequence 15, Appl

ALIGNMENTS

```
RESULT 1
US-09-140-804-5
; Sequence 5, Application US/09140804
; Patent No. 6197930
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
; FILE REFERENCE: 97-49
; CURRENT APPLICATION NUMBER: US/09/140,804
; EARLIER FILING DATE: 1998-08-26
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Tamias sibiricus
US-09-140-804-5

Query Match      8.9%: Score 221; DB 4; Length 215;
Best Local Similarity 34.8%; Pred. No. 7.3e-14;
Matches 57; Conservative 20; Mismatches 69; Indels 18; Gaps 3;

QY 16 EVAGPPAHRPP-----EEVGPPAGALGQYTGTEISEMTKCPDIERSATFV 63
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 37 EPGCGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 94
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 64 KISGKLPDPKPIITFGVLYNMRDLKEMGVFACVPCGVYSSFVEVLEHCKVIMLR 123
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 95 KLSRPPPEFPQPIVFKEALYNOGSHFNMATGERSCVLPGVYNGFDIRLFOSSVKIRLMR 154
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 124 KOILANKKEI-----SKQSIQEVTVWLKAFSTRIRAEKSSFN 163
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 155 DGIQVREKQANDSYKHAMGSYIMALGKDKVWLESKLGTES 198
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 2
US-09-140-804-6
; Sequence 6, Application US/09140804
; Patent No. 6197930
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
; FILE REFERENCE: 97-49
; CURRENT APPLICATION NUMBER: US/09/140,804
; EARLIER FILING DATE: 1998-08-26
```

EARLIER APPLICATION NUMBER: 60/056,983
EARLIER FILING DATE: 1997-08-26
NUMBER OF SEQ ID NOS: 47
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO: 6
LENGTH: 236
TYPE: PRT
ORGANISM: Tamias sibiricus
US-09-140-804-6

Query Match 7.0%; Score 174; DB 4; Length 236;
Best Local Similarity 31.3%; Pred. No. 3.7e-09;
Matches 47; Conservative 15; Mismatches 74; Indels 14; Gaps 2;

QY 6 VLTAVIEHVEVAGPPAHPRPEEVEVPPGA-----GDLPOYTGISEMTKCPG 53
DB 48 ILALSVLMHVCSNVPGQPGPMGPRGPGKPGPPGGLGPPGPMYVNCNS 107
QY 54 PDIERSAFTVKSGLPLPFKPIIFTGVLYNAQRDLKEMGVACRYPGNYSSPDVELH 113
DB 108 KG--TSAFAVKANELPPAPSQVIFKEALHDQGHFDLATGVFTCPGLYQFGPHIEAV 165
QY 114 HCKVNIWMRKQIILANKEEISKOOSIOEVT 143
DB 166 QRAVKYSLMRNGTOVMREARQAQDYEHTS 195

RESULT 3
US-08-463-911-3
Sequence 3, Application US/08463911
Patent No. 5869330
GENERAL INFORMATION:
APPLICANT: Scherer, Philipp E.
APPLICANT: Lodish, Harvey F.
TITLE OF INVENTION: A NOVEL SERUM PROTEIN PRODUCED
TITLE OF INVENTION: EXCLUSIVELY IN ADIPOCYTES
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESSES:
ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,911
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: WH195-05
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 185 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-463-911-3

Query Match 6.7%; Score 167; DB 2; Length 185;
Best Local Similarity 35.2%; Pred. No. 1.2e-08;

Matches 45; Conservative 11; Mismatches 64; Indels 8; Gaps 3;
QY 19 GPPAHPRPEEYVPP---GAGLPDYTGISEMTKCPDIERSAFTVKSGLPLPFKP 75
DB 22 GPGTPPKRP---GPPGMNGFPLGPPGPMYVNCNSKG--TSAFAVKANELPPAPSQP 76
QY 76 IFTGVLYNAQRDLKEMGVACRYPGNYSSPDVELHCKVNIWMRKQIILANKEEISK 135
DB 77 VIFKEALHDQGHFDLATGVFTCPGLYQFGPHIEAVQRAVKYSLMRNGTOVMREAEA 136
QY 136 QOSIOEVT 143
DB 137 QDYEHTS 144

RESULT 4
US-08-463-911-2
Sequence 2, Application US/08463911
Patent No. 5869330
GENERAL INFORMATION:
APPLICANT: Scherer, Philipp E.
APPLICANT: Lodish, Harvey F.
TITLE OF INVENTION: A NOVEL SERUM PROTEIN PRODUCED
TITLE OF INVENTION: EXCLUSIVELY IN ADIPOCYTES
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESSES:
ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,911
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: WH195-05
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 247 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-463-911-2

Query Match 6.1%; Score 152.5; DB 2; Length 247;
Best Local Similarity 29.7%; Pred. No. 5.3e-07;
Matches 44; Conservative 23; Mismatches 64; Indels 17; Gaps 4;

QY 31 GPPAGBLPYTGISEMTKCPDIERSAFTVKSGLPLPFKPIIFTGVLYNAQRDLK 90
DB 93 GPGFPGTPPKRGEPEEAY----MYRSAFSVGLFRTVTPNVPIRFTKIFYNQONHYD 147
QY 91 EAMGVACRYPGNYSSPDVELH--HCKVNIWMRKQIILANKEEISKOOSIOEVTWLLK 148
DB 148 GSTGKFCYNIRPLGYFSYHITVYKDKVSLFKKRAVLETTYDQYQERNVDQASGSYLLH 207
QY 149 A-----FSFIREAEKSSENLHPDNV 169
DB 208 LEVGDQWMLQYRGSDH---NGLYADNV 232

```

RESULT 5
US-08-463-911-7
; Sequence 7, Application US/08463911
; Patent No. 5869330
; GENERAL INFORMATION:
; APPLICANT: Scherer, Philipp E.
; APPLICANT: Lodish, Harvey F.
; TITLE OF INVENTION: A NOVEL SERUM PROTEIN PRODUCED
; TITLE OF INVENTION: EXCLUSIVELY IN ADIPOCYTES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESSES:
; ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Millitia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/463,911
; APPLICATION NUMBER: US/08/463,911
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WHI95-05
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 244 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-463-911-7

Query Match 6.0%; Score 148.5; DB 2; Length 244;
Best Local Similarity 30.7%; Pred. No. 1.3e-06;
Matches 43; Conservative 16; Mismatches 48; Indels 33; Gaps 4;

QY 17 VAGPPAHP-----RPEEVGPPGAPGLPQYTGSEISMTCPC-----CPDIE 57
DB 40 MAGIGFCHGNAPGRDGDGTPEGEKGDPLIGPKDIDG-TGVPAEGPRGPGIO 98
QY 58 -----RSAFYTKLSGKLPFPKPIFTGVLYNAORDLKEAMGVFACRYPGNY 104
DB 99 GKRGEPGEGAYVRSASFVSGLETYYTIPNMPIRFTKIFYNQNHNDGSGTKFHCNIPGLY 158
QY 105 YSFVEVLHCKVNIWLMRK 124
DB 159 YFAYHITYMKDKVSLFKK 178

RESULT 6
US-09-140-804-3
; Sequence 3, Application US/09140804
; Patent No. 6197930
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Humes, Jacqueline M.
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
; FILE REFERENCE: 97-49
; CURRENT APPLICATION NUMBER: US/09/140,804
; CURRENT FILING DATE: 1998-08-26
; EARLIER APPLICATION NUMBER: 60/056,983
; EARLIER FILING DATE: 1997-08-26

```

```

; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-140-804-3

Query Match 6.0%; Score 148.5; DB 4; Length 244;
Best Local Similarity 30.7%; Pred. No. 1.3e-06;
Matches 43; Conservative 16; Mismatches 48; Indels 33; Gaps 4;

QY 17 VAGPPAHP-----RPEEVGPPGAPGLPQYTGSEISMTCPC-----CPDIE 57
DB 40 MAGIGFCHGNAPGRDGDGTPEGEKGDPLIGPKDIDG-TGVPAEGPRGPGIO 98
QY 58 -----RSAFYTKLSGKLPFPKPIFTGVLYNAORDLKEAMGVFACRYPGNY 104
DB 99 GKRGEPGEGAYVRSASFVSGLETYYTIPNMPIRFTKIFYNQNHNDGSGTKFHCNIPGLY 158
QY 105 YSFVEVLHCKVNIWLMRK 124
DB 159 YFAYHITYMKDKVSLFKK 178

RESULT 7
US-09-140-804-8
; Sequence 8, Application US/09140804
; Patent No. 6197930
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Humes, Jacqueline M.
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
; FILE REFERENCE: 97-49
; CURRENT APPLICATION NUMBER: US/09/140,804
; CURRENT FILING DATE: 1998-08-26
; EARLIER APPLICATION NUMBER: 60/056,983
; EARLIER FILING DATE: 1997-08-26
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 247
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-140-804-8

Query Match 5.8%; Score 144.5; DB 4; Length 247;
Best Local Similarity 26.4%; Pred. No. 3.3e-06;
Matches 51; Conservative 25; Mismatches 74; Indels 43; Gaps 6;

QY 17 VAGPPAHP-----RPEEVGPPGAPGLPQYTGSEISMTCPCPD----- 55
DB 43 MAGIGFCHGNAPGRDGDGTPEGEKGDPLIGPKDIDG-TGVPAEGPRGPGIO 102
QY 56 -----IRSAFTYKLSGKLPFPKPIFTGVLYNAORDLKEAMGVFACRYPGNY 105
DB 103 RKGEPEGALMYRVSASFVSGLETYYTIPNMPIRFTKIFYNQNHNDGSGTKFHCNIPGLY 162
QY 106 SSFVEVLH-HCKVNIWLMRKQILANKKEISKQOSIQEVTWVLLA-----FSFIREA 156
DB 163 FSYHITYMKDKVSLFKKDKAVLFTYDQYQEKKNVDAQSGVLLHLEVDQWMLQVYGDG 222
QY 157 EHKSSENLHPDNV 169
DB 223 DH---NGLYADNV 232

RESULT 8
US-09-118-408-3
; Sequence 3, Application US/09118408A
; Patent No. 6265544

```

```

: GENERAL INFORMATION:
: APPLICANT: Shepard, Paul O.
: TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
: FILE REFERENCE: 97-30
: CURRENT APPLICATION NUMBER: US/09/118,408A
: CURRENT FILING DATE: 1998-07-17
: EARLIER APPLICATION NUMBER: 60/053,154
: EARLIER FILING DATE: 1997-07-18
: NUMBER OF SEQ ID NOS: 47
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 3
: LENGTH: 247
: TYPE: PRT
: ORGANISM: Homo sapien
US-09-118-408-3

Query Match          5.8%; Score 144.5; DB 4; Length 247;
Best Local Similarity 26.4%; Pred. No. 3.3e-06;
Matches 51; Conservative 25; Mismatches 74; Indels 43; Gaps 6;

QY 17 VAGPPAHP-----RPEEVGPPGAPGLPYTGISE--MTKCPD-----55
   || || || || || || || || || || || || || || || || ||
DB 43 MAGIPGPHNGPGRDGRDGPGEKKGDAGLGPKGETGVGMGAEGRGFQTPG 102
QY 56 -----IERSAFYVLSGKLPPEKPIITFTGVLYNAORDLKEAMGVFACRVPGNTY 105
   : || || : : : || || : : : || || : : : || || : : : ||
DB 103 RKEPEEAAYMYRSASFVGLFTRVTPVPIRFTRKIFYNQONHDSGTGFYCNIPGLY 162
QY 106 SPSDEVEH--HCKVNIWLMRKQILANKKEISKQOSIQEIVWVLLKA-----FSFIREA 156
   | : : : || || : : : || || : : : || || : : : || ||
DB 163 FSHITVYMKDVKVSLEFKDKAVLFTTYDQYOEKNVDAASSGLHLEVGDDQWLQYTGDC 222
QY 157 EHKSENLPDNY 169
   : | : | : |
DB 223 DH---NGLYADNY 232

RESULT 9
US-09-188-930-295
: Sequence 295, Application US/09188930A
: Patent No. 6150502
: GENERAL INFORMATION:
: APPLICANT: Watson, James D.
: APPLICANT: Strachan, Lorna
: APPLICANT: Sleeman, Matthew
: APPLICANT: Onrust, Rene
: APPLICANT: Murison, James Greg
: TITLE OF INVENTION: Compositions Isolated From Skin Cells
: FILE REFERENCE: 11000.1011c1
: CURRENT APPLICATION NUMBER: US/09/188,930A
: CURRENT FILING DATE: 1998-11-09
: NUMBER OF SEQ ID NOS: 348
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 295
: LENGTH: 243
: TYPE: PRT
: ORGANISM: Rat
US-09-188-930-295

Query Match          5.0%; Score 124; DB 4; Length 243;
Best Local Similarity 31.9%; Pred. No. 0.00034;
Matches 38; Conservative 12; Mismatches 53; Indels 16; Gaps 4;

QY 19 GPPAHPPEEVGPPGAPGLPYTGE-----ISEMTKCCPDIERSAFTYKLS 66
   | : | : | : | : | : | : | : | : | : | : | : | : |
DB 54 GRDGAAGAGEEGEGRPLPGPRGEPGRGAGPVGAISPAGECVP--PRSAFSAKRS 111
QY 67 -GKLPLPF-KPIITFTGVLYNAORDLKEAMGVFACRVPGNTYSSFDVELHCKVNIWLM 123
   : | : | : | : | : | : | : | : | : | : | : | : |
DB 112 ESNVPPADPLPFPDRLVNLNEGHDATITGKFTQVGVYFAVHATVYRASLQFDLVK 170
```

```

RESULT 10
US-09-188-930-294
: Sequence 294, Application US/09188930A
: Patent No. 6150502
: GENERAL INFORMATION:
: APPLICANT: Watson, James D.
: APPLICANT: Strachan, Lorna
: APPLICANT: Sleeman, Matthew
: APPLICANT: Onrust, Rene
: APPLICANT: Murison, James Greg
: TITLE OF INVENTION: Compositions Isolated From Skin Cells
: FILE REFERENCE: 11000.1011c1
: CURRENT APPLICATION NUMBER: US/09/188,930A
: CURRENT FILING DATE: 1998-11-09
: NUMBER OF SEQ ID NOS: 348
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 294
: LENGTH: 294
: TYPE: PRT
: ORGANISM: Rat
US-09-188-930-294

Query Match          5.0%; Score 123.5; DB 4; Length 294;
Best Local Similarity 31.8%; Pred. No. 0.00051;
Matches 34; Conservative 14; Mismatches 52; Indels 7; Gaps 2;

QY 17 VAGPPAHPPEEVGPPGAPGLPYTGISEMTKCPDIE-NSAFYVLSGKLPPEKPP 75
   | : | : | : | : | : | : | : | : | : | : | : | : |
DB 122 VSGTPGKHGIPKKGPKKGKGEPLPG-----PCSGSSRAKSAVSATVTSYPERRLP 175
QY 76 IIFTGVLYNAORDLKEAMGVFACRVPGNTYSSFDVELHCKVNIWLM 122
   | : | : | : | : | : | : | : | : | : | : | : | : |
DB 176 IKFDKILNMGSHYNASSGKFTVCSPGIYFTYDITLANKHLAIGLV 222
```

```

RESULT 11
US-09-140-804-2
: Sequence 2, Application US/09140804
: Patent No. 6197930
: GENERAL INFORMATION:
: APPLICANT: Shepard, Paul O.
: APPLICANT: Humes, Jacqueline M.
: TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
: FILE REFERENCE: 97-49
: CURRENT APPLICATION NUMBER: US/09/140,804
: CURRENT FILING DATE: 1998-08-26
: EARLIER APPLICATION NUMBER: 60/056,983
: EARLIER FILING DATE: 1997-08-26
: NUMBER OF SEQ ID NOS: 47
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 2
: LENGTH: 243
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-140-804-2

Query Match          4.9%; Score 121.5; DB 4; Length 243;
Best Local Similarity 27.5%; Pred. No. 0.00059;
Matches 42; Conservative 15; Mismatches 57; Indels 39; Gaps 6;

QY 14 HVEVAGPPAH-----PRPPEVGPAPGLPYTGE-----44
   | : | : | : | : | : | : | : | : | : | : | : | : |
DB 31 HPGLETPRGHSGSLGPRDGRDGDGAPGAEKGGGPGLPGRGDPGRGEGPAG 90
QY 45 -ISEMTKCPDIERSAFTYKLS-GKLPLPF-KPIITFTGVLYNAORDLKEAMGVFACRV 101
   : | : | : | : | : | : | : | : | : | : | : | : |
DB 91 PTPGAGECVP--PRSAFSAKRSRVPFPPSDAPLPFPDRLVNLNEGHDATVYGTQCV 148
```

QY 102 GNYSSFDVLEHCKVNTLMARKOILLANKEEIS 134
 DB 149 GYYEFAVHATYVRASLOQFDLVK-----NGESIA 176

RESULT 12
 US-09-485-316A-12

Sequence 12, Application US/09485316A
 Patent No. 6344441
 GENERAL INFORMATION:
 APPLICANT: Bihain, Bernard
 APPLICANT: Bougueleret, Lydie
 APPLICANT: Yen-Potin, Frances
 TITLE OF INVENTION: Lipoprotein-regulating medicaments
 FILE REFERENCE: GENSET.036APC
 CURRENT APPLICATION NUMBER: US/09/485,316A
 PRIOR FILING DATE: 2000-02-04
 PRIOR APPLICATION NUMBER: FR 97/10088
 PRIOR FILING DATE: 1997-08-06
 PRIOR APPLICATION NUMBER: FR 98/05032
 PRIOR FILING DATE: 1998-04-22
 PRIOR APPLICATION NUMBER: PCT IB98/01256
 PRIOR FILING DATE: 1998-08-06
 NUMBER OF SEQ ID NOS: 14
 SOFTWARE: Patent.pm
 SEQ ID NO 12
 LENGTH: 130
 TYPE: PRT
 ORGANISM: Mus musculus
 FEATURE:
 NAME/KEY: SIMILAR
 LOCATION: 1..130
 OTHER INFORMATION: fragment 118..247 of translation from ref Genbank U49915
 US-09-485-316A-12

Query Match 4.7%; Score 116; DB 4; Length 130;
 Best Local Similarity 27.1%; Pred. No. 0.00079;

Matches 32; Conservative 22; Mismatches 52; Indels 12; Gaps 3;

QY 61 FTVKLSGKLPFPKPIIFGVLYNAQRDLKAMGVACRPNGYSSFDVELH--HCKYN 118
 DB 1 FSVGLETRVTPVNPPIRFKIFYNQONHNDSTGKRYCNPGLYPSYHITVYMKDVKS 60
 QY 119 IWLARKOILLANKEEISKOOSIOEVTWVLLKA-----FSFIREAEKSENLHPDNV 169
 DB 61 LFKKDAVLEFTYDYOEKNVDAQSGSVLLHLEVGDOVWLQVYGDGH---NGLYADNV 115

RESULT 13
 US-09-485-316A-13

Sequence 13, Application US/09485316A
 Patent No. 6344441
 GENERAL INFORMATION:
 APPLICANT: Bihain, Bernard
 APPLICANT: Bougueleret, Lydie
 APPLICANT: Yen-Potin, Frances
 TITLE OF INVENTION: Lipoprotein-regulating medicaments
 FILE REFERENCE: GENSET.036APC
 CURRENT APPLICATION NUMBER: US/09/485,316A
 PRIOR FILING DATE: 2000-02-04
 PRIOR APPLICATION NUMBER: FR 97/10088
 PRIOR FILING DATE: 1997-08-06
 PRIOR APPLICATION NUMBER: FR 98/05032
 PRIOR FILING DATE: 1998-04-22
 PRIOR APPLICATION NUMBER: PCT IB98/01256
 PRIOR FILING DATE: 1998-08-06
 NUMBER OF SEQ ID NOS: 14
 SOFTWARE: Patent.pm
 SEQ ID NO 13
 LENGTH: 130
 TYPE: PRT
 ORGANISM: Mus musculus

FEATURE:
 NAME/KEY: SIMILAR
 LOCATION: 1..130
 OTHER INFORMATION: fragment 118..267 of translation from ref genbank U37222
 US-09-485-316A-13

Query Match 4.6%; Score 114; DB 4; Length 130;
 Best Local Similarity 27.1%; Pred. No. 0.0012;
 Matches 32; Conservative 22; Mismatches 52; Indels 12; Gaps 3;

QY 61 FTVKLSGKLPFPKPIIFGVLYNAQRDLKAMGVACRPNGYSSFDVELH--HCKYN 118
 DB 1 FSVGLETRVTPVNPPIRFKIFYNQONHNDSTGKRYCNPGLYPSYHITVYMKDVKS 60
 QY 119 IWLARKOILLANKEEISKOOSIOEVTWVLLKA-----FSFIREAEKSENLHPDNV 169
 DB 61 LFKKDAVLEFTYDYOEKNVDAQSGSVLLHLEVGDOVWLQVYGDGH---NGLYADNV 115

RESULT 14

US-08-328-254-6
 Sequence 6, Application US/08328254
 Patent No. 5710022

GENERAL INFORMATION:
 APPLICANT: Zhu, Xueliang
 APPLICANT: Lee, Wen-Hwa
 TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Campbell and Flores
 STREET: 4370 La Jolla Village Drive, Suite 700
 CITY: San Diego
 STATE: California
 COUNTRY: USA
 ZIP: 92132

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/328,254
 FILING DATE: 24-OCT-1994
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/141,239
 FILING DATE: 22-OCT-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Campbell, Cathryn A.
 REGISTRATION NUMBER: 31,815
 REFERENCE/DOCKET NUMBER: P-CJ 1191
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 535-9001
 TELEFAX: (619) 535-8949
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2482 amino acids
 TYPE: amino acid
 TOPOLOGY: linear

MOLECULE TYPE: protein
 US-08-328-254-6

Query Match 4.4%; Score 110; DB 1; Length 2482;
 Best Local Similarity 21.4%; Pred. No. 0.3; Indels 112; Gaps 17;
 Matches 80; Conservative 52; Mismatches 129;

QY 116 KYNIMLMARKOILLANKEEIKOO-----SIOEVTWVLLKAFFSFIREAHSSENLHPD 167
 DB 2090 EYEIQYRRKLSKESCSQKLEIDLLKSSKEELNLSIKATTOILEELAKTKRMKNLKV 2149
 QY 168 NVIKKKNPSEGRFKLAIEIC-----ICNEELN---VNPQDNG----- 202

Db 2150 NOLKKENERAOGKMKLLIKSCQLEKEKEILLQKELSQLQAQOEKOKTGYMDTKVDELTT 2209
QY 203 --ENISWTCORSSQOS-----IKSLAMRPRRMFCGTCGSLCCVQPRD----- 244
Db 2210 EIKELKTELEKTEKDEYLDKCYSLLSHEKLEKAKEMLETQVAHLCSQOSKODSRGSP 2269
QY 245 -LVPCVYVNSAVAASEGASPKP---WOLPSGVEPVGAKKSRIE-VWE-----PIPRF 290
Db 2270 LLGPVVP-----GPSPIPSYTEKRLSSGQNKASGKRQSSGIMENCGPTPATPESF 2321
QY 291 QKIYGNPMNPRQKFAV--GVGSSWRTSA-----RVQKGNGWEPHRRVPSGAPSS 339
Db 2322 SK-----KSKKAVMSGIHPAEDTEGTEFEPEGLPEVVKKGFPAD-----IPYKTS 2367
QY 340 RAVRSPSSRLQKGRSTSL-----QHYPEKSTDT---OCQPYKAGMESVPYK 386
Db 2368 YILRTTMTATSPRLAOKLALSPLSLCKENLAESSKPTAGSGRSQKVAQRSFVDSG 2427
QY 387 TVVAE-LTKTVGI 398
Db 2428 TILREPTTKSVPV 2440

RESULT 15
US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCORE PROTEIN,
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DORPMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DDS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYDROTHERMAL: ND
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
; US-08-353-700-1

Query Match 4.4%; Score 110; DB 1; Length 3248;
Best Local Similarity 21.4%; Pred. No. 0.45;

Matches 80; Conservative 52; Mismatches 129; Indels 112; Gaps 17;
QY 116 KVINIMLRKOILANKKEESTKQO-----SIOBYTVLLKAFSFRFAEHKSSENLHPD 167
Db 2818 EVELQYRREKLKSKKECCSSQKLEIDLKSKKEBELNLSKATQIILELKKTKMNLKYY 2877
QY 168 NVIKKKNPSECKFLAAEIC-----ICNEELN-----VNPQDN----- 202
Db 2878 NOLKKENERAOGKMKLLIKSCQLEKEKEILLQKELSQLQAQOEKOKTGYMDTKVDELTT 2937
QY 203 --ENISWTCORSSQOS-----IKSLAMRPRRMFCGTCGSLCCVQPRD----- 244
Db 2938 EIKELKTELEKTEKDEYLDKCYSLLSHEKLEKAKEMLETQVAHLCSQOSKODSRGSP 2997
QY 245 -LVPCVYVNSAVAASEGASPKP---WOLPSGVEPVGAKKSRIE-VWE-----PIPRF 290
Db 2998 LLGPVVP-----GPSPIPSYTEKRLSSGQNKASGKRQSSGIMENCGPTPATPESF 3049
QY 291 QKIYGNPMNPRQKFAV--GVGSSWRTSA-----RVQKGNGWEPHRRVPSGAPSS 339
Db 3050 SK-----KSKKAVMSGIHPAEDTEGTEFEPEGLPEVVKKGFPAD-----IPYKTS 3095
QY 340 RAVRSPSSRLQKGRSTSL-----QHYPEKSTDT---OCQPYKAGMESVPYK 386
Db 3096 YILRTTMTATSPRLAOKLALSPLSLCKENLAESSKPTAGSGRSQKVAQRSFVDSG 3155
QY 387 TVVAE-LTKTVGI 398
Db 3156 TILREPTTKSVPV 3168

Search completed: June 30, 2002, 11:53:23
Job time: 144 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2002, 11:49:04 ; Search time 31.16 Seconds

(without alignments)
1636.164 Million cell updates/sec

Title: US-09-997-610-2

Perfect score: 2494

Sequence: 1 IVVLPVLTAVIEHVEVACP.....GQFPFGTAVTQCLTLCRM 459

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	766	30.7	332	22	ABG02620
2	698	28.0	1984	22	ABG01338
3	694.5	27.8	273	22	ABG00491
4	676.5	27.1	939	22	ABG21982
5	665.5	26.7	901	22	ABG16683
6	653	26.2	220	22	ABG2981
7	628.5	25.2	333	22	ABG02511
8	602	24.1	466	22	ABG16460
9	573.5	23.0	270	22	ABG00481
10	570	22.9	439	22	ABG16421
11	570	22.9	439	22	ABG17512

12	540.5	21.7	261	22	ABG02502	Novel human diagno
13	540.5	21.7	261	22	ABG02632	Novel human diagno
14	528	21.2	766	22	ABG17044	Novel human diagno
15	523	21.0	636	22	ABG27632	Novel human diagno
16	487	19.5	561	22	ABG17236	Novel human diagno
17	486	19.5	290	22	ABG03735	Novel human diagno
18	483	19.4	595	22	ABG11326	Novel human diagno
19	477.5	19.1	239	22	ABG08795	Novel human diagno
20	476	19.1	533	22	ABG00280	Novel human diagno
21	475.5	19.1	310	22	ABG16947	Novel human diagno
22	473	19.0	224	22	ABG02622	Novel human diagno
23	462	18.5	289	22	ABG07801	Novel human diagno
24	459.5	18.4	553	22	ABG04234	Novel human diagno
25	458.5	18.3	297	22	ABG26263	Novel human diagno
26	457.5	18.3	842	22	ABG05486	Novel human diagno
27	435.5	17.5	528	22	ABG29732	Novel human diagno
28	434.5	17.4	276	22	ABG17722	Novel human diagno
29	424	17.0	204	22	ABG00278	Novel human diagno
30	417.5	16.7	1218	22	ABG06054	Novel human diagno
31	417.5	16.7	1218	22	ABG09637	Novel human diagno
32	417.5	16.7	1218	22	ABG10096	Novel human diagno
33	417.5	16.7	1218	22	ABG10796	Novel human diagno
34	417.5	16.7	1218	22	ABG14538	Novel human diagno
35	417.5	16.7	1218	22	ABG14753	Novel human diagno
36	417.5	16.7	1218	22	ABG17474	Novel human diagno
37	417.5	16.7	1218	22	ABG19905	Novel human diagno
38	417.5	16.7	1223	22	ABG08695	Novel human diagno
39	408.5	16.4	371	22	ABG05384	Novel human diagno
40	395	15.8	343	22	ABG16673	Novel human diagno
41	393	15.8	258	22	ABG00713	Novel human diagno
42	393	15.8	665	22	ABG03829	Novel human diagno
43	387	15.5	416	22	ABG17252	Novel human diagno
44	385	15.4	486	22	ABG11294	Novel human diagno
45	374.5	15.0	255	22	ABG17728	Novel human diagno

ALIGNMENTS

RESULT 1
ID ABG02620 standard; Protein: 332 AA.
AC ABG02620;
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #2611.
XX
KW Human; chromosome mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PM W0200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
XX
DR N-PSDB: AAS66807.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess

Db	1766	pnhvtpgcalpsgtvrrrpgssrrpnpqgriskshhcbpgkasdqtgshpmkaatgpsaag1	1825
Qy	384	-----PYKTVVAELRKTV 396	
Db	1826	lefargplgtlfawvtaav 1844	
RESULT	3		
ABG00491			
ID	ABG00491	standard; Protein; 273 AA.	
XX	AC	ABG00491;	
XX	DT	13-FEB-2002 (first entry)	
XX	DE	Novel human diagnostic protein #482.	
XX	KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;	
XX	KW	food supplement; medical imaging; diagnostic; genetic disorder.	
XX	OS	Homo sapiens.	
XX	PN	WO200175067-A2.	
XX	PD	11-OCT-2001.	
XX	PF	30-MAR-2001; 2001WO-US08631.	
XX	PR	31-MAR-2000; 2000US-0540217.	
XX	PR	23-AUG-2000; 2000US-0649167.	
XX	PA	(HYSE-) HYSEQ INC.	
XX	PI	Dmanac RT, Liu C, Tang YF;	
XX	DR	WPI: 2001-639362/73.	
XX	DR	N-PSDB: AAS64678.	
XX	PT	New isolated polynucleotide and encoded polypeptides, useful in	
XX	PT	diagnostics, forensics, gene mapping, identification of mutations	
XX	PT	responsible for genetic disorders or other traits and to assess	
XX	PT	biodiversity -	
XX	XX		
XX	PS	Claim 20: SEQ ID No 30850; 103bp; English.	
XX	CC	The invention relates to isolated polynucleotide (I) and	
XX	CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,	
XX	CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome	
XX	CC	and gene mapping, and in recombinant production of (II). The	
XX	CC	polynucleotides are also used in diagnostics as expressed sequence tags	
XX	CC	for identifying expressed genes. (I) is useful in gene therapy techniques	
XX	CC	to restore normal activity of (II) or to treat disease states involving	
XX	CC	(II). (II) is useful for generating antibodies against it, detecting or	
XX	CC	quantitating a polypeptide in tissue, as molecular weight markers and as	
XX	CC	a food supplement. (II) and its binding partners are useful in medical	
XX	CC	imaging of sites expressing (II). (I) and (II) are useful for treating	
XX	CC	disorders involving aberrant protein expression or biological activity.	
XX	CC	The polypeptide and polynucleotide sequences have applications in	
XX	CC	diagnostics, forensics, gene mapping, identification of mutations	
XX	CC	responsible for genetic disorders or other traits to assess biodiversity	
XX	CC	and to produce other types of data and products dependent on DNA and	
XX	CC	amino acid sequences. ABG00010-ABG30377 represent novel human	
XX	CC	diagnostic amino acid sequences of the invention.	
XX	CC	Note: The sequence data for this patent did not appear in the printed	
XX	CC	specification, but was obtained in electronic format directly from WIPO	
XX	CC	at ftp.wipo.int/pub/published_pct_sequences .	
XX	XX		
XX	XX	Sequence 273 AA;	
XX	XX		
Query Match	27.8%	Score 694.5, DB 22, Length 273,	
Best Local Similarity	58.5%	Pred No. 6, 8e-53;	

	Matches	141	Conservative	27	Mismatches	56	Indels	17	Gaps
QY	110	VELHHCKNINW--LMRKOILANKKEISIKSOOSIQEYTWVWLKAFSFRREAHKSSSEMLHP	166						
Db	28	lelgnkqgrdlmselerdrlryleeelfkqgsigemlwllkafsfvrgeahkslenlqp	87						
QY	167	DNVIRKKNNFSGSKFKFLAEICICNSELWVNPDPDNGENISWTCORSSQOOSIKSLAMPRR	226						
Db	88	dsavekklpfiseekfkfpaaelcvssseepnvkpgenrenfraqckasqgvpvsqaqprtr	147						
QY	227	-KMFCEGTGSGSLCCVQPRDLVPCVPVNS-----AVASEGASPKPMQLPSGVE	272						
Db	148	kkvfclgldeslccvlsrldlvpcllaasapanwkrqgdfcfdgmaassgaahkphlpgoye	207						
QY	273	PVGAKRSRIEWEPPILRFQKIYGNPMPPRQKFAVGVGSSWRTSRARVYQKGNWGVEPPHRV	332						
Db	208	pmgldksgqtevwepjprfdgmyknaawmsrqkfaagaepepswrtisaravlkgnvgspeptgs	267						
QY	333	P	333						
Db	268	P	268						
RESULT 4									
ABG21982	ID	ABG21982	standard; protein; 939 AA.						
AC	XX	ABG21982;							
DT	XX	18-FEB-2002	(first entry)						
DE	XX	Novel human diagnostic protein #21973.							
KW	XX	Human; chromosome mapping; gene mapping; gene therapy; forensic;							
KW	XX	food supplement; medical imaging; diagnostic; genetic disorder.							
OS	XX	Homo sapiens.							
PN	XX	WO200175067-A2.							
PD	XX	11-OCT-2001.							
PF	XX	30-MAR-2001; 2001WO-US08631.							
PR	XX	31-MAR-2000; 2000US-0540217.							
PR	XX	23-AUG-2000; 2000US-0649167.							
PA	XX	(HYSE-) HYSEQ INC.							
PI	XX	Dmanac RT, Liu C, Tang YT;							
DR	XX	WPI: 2001-6399362/73.							
DR	XX	N-PSDB; AAS86169.							
PT	XX	New isolated polynucleotide and encoded polypeptides, useful in							
PT	XX	diagnostics, forensics, gene mapping, identification of mutations							
PT	XX	responsible for genetic disorders or other traits and to assess							
PT	XX	biodiversity							
PS	XX	Claim 20; SEQ ID No 52341; 103pp; English.							
CC	XX	The invention relates to isolated polynucleotide (I) and							
CC	XX	polypeptide (II) sequences. (I) is useful as hybridisation probes,							
CC	XX	polymerase chain reaction (PCR) primers, oligomers, and for chromosome							
CC	XX	and genome mapping, and in recombinant production of (II). The							
CC	XX	polynucleotides are also used in diagnostics as expressed sequence tags							
CC	XX	for identifying expressed genes. (I) is useful in gene therapy techniques							
CC	XX	to restore normal activity of (II) or to treat disease states involving							
CC	XX	(II). (II) is useful for generating antibodies against it, detecting or							
CC	XX	quantitating a polypeptide in tissue, as molecular weight markers and as							
CC	XX	a food supplement. (II) and its binding partners are useful in medical							
CC	XX	imaging of sites expressing (II). (I) and (II) are useful for treating							
CC	XX	disorders involving aberrant protein expression or biological activity.							

CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

Sequence 939 AA:

Query Match 27.1%; Score 676.5; DB 22; Length 939;

Best Local Similarity 39.7%; Pred. No. 1.5e-50;

Matches 184; Conservative 35; Mismatches 101; Indels 143; Gaps 16;

QY 2 VVPLVLTAVIEHVEYACPPAP--RPEEYGPFG--APGLPOTYGTGISMTCPCPDIE 57
DB 79 lvspl-----hfrvggkpaapavrsqgvgkvgnvsnfr----- 114
QY 58 RSAPVTKLSGKLPPEKPIIFGVLYNQAQDLKAMGVACRVPGNY----- 105
DB 115 --lflekltqifndyfpflmagnefnlfvd-----kygdlrptsltsa 158
QY 106 -----SSEFDV---ELHCKVNIWLMKQ---ILAN-----KEE 132
DB 159 sdncqjmalarfdaavgrllypcpkkqlidkeigjllahprdlwnfkfegedlgylaee 218
QY 133 ISKQSIQIOWTWLTKAFSEITREAHKSSNLHPDNVTKKKNPSEKFLAEITCME 192
DB 219 mfyqgtcgevtwllkatsfkretkhsseellqpd-vierknpsiectlilaecisnuk 277
QY 193 ELNVNPDQNGENISWTCQSSQSIKSLAMRRKRFCTGPGSLCCVQPRDLVPCVPV 252
DB 278 alvnhqddngenvsracqpfqll-----qpw-----In 307
QY 253 SAVASGASGAPKPV-----QLPSG---VEPVGAKKSRIEWEPPIRFKTYGNPMPKROK 303
DB 308 gakvqjql----wlgtrgalshgnfhmvgpygaeksrlwkpilpgfqmrygkamrsgk 363
QY 304 FAVGVSSWMTSARVYOKGNWGMEDPHRVPSGAPSSRAVRSSPPSSRLQGRSTDSLQHV 363
DB 364 faag-----avqkgnvyleshtrvcpappsgavtrrgplsrpqngrtsdshra 413
QY 364 PEKSTDTQCQPYKAAGMESVPYKTVAAELTKTVGITYLLHCHDL 406
DB 414 pykadtceqpmksasesetvpckatrvelpkrmtgahllrhd1 456

RESULT 5

ABG16683 ID ABG16683 standard; Protein; 901 AA.

XX ABG16683:

XX 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #16674.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0648167.

XX (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

DR N-PSDB; AAS80870.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

Claim 20: SEQ ID No 47042; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX

Sequence 901 AA:

Query Match 26.7%; Score 665.5; DB 22; Length 901;

Best Local Similarity 44.3%; Pred. No. 1.3e-49;

Matches 162; Conservative 31; Mismatches 88; Indels 85; Gaps 12;

QY 128 ANKEEISKOQSIQEWYTWLTKAF--SFTREAHKSSNLHPDNVTKKKNPSE--GKF- 181
DB 576 afksndsgqdvsetfthklldwledafgmkaeetdee-----kxpmwmlfygrfl 628
QY 182 -----KLAA---EICTCNEELNVNPDQNGENISWTCQSSQSIKSLAMRRKRW 228
DB 629 avvglegtsmaatrenev-----evkvetqdkpflfiw----- 662
QY 229 FCGTGG-----SLCCVQPRDLVPCVPVNSA-----VASGASGAPKMPOLPS 269
DB 663 -----ghlfyfyllssccchprnlvpcvsaapafarrgqvarplvsasppkphlpr 716
QY 270 GVEPVGAKKSRIEWEPPIRFKTYGNPMPKROKFAVGSSWMTSARVYOKGNWGMEDP 329
DB 717 gvepagaqgsirevwapprifgkmyrnaamrsgkfalelssvtaravtrkrnyglpct 776
QY 330 HRVPSGAPSSRAVRSSPPSSRLQGRSTDSLQHVPEKSTDTQCQPYKAAGMESVPYKTV 389
DB 777 qrvptgappsgvtvrrpplsrspgnvrtscdshcvygaadtqbpkaarraevpckatg 836
QY 390 AELTKTVGITYLLHCHDLVRRHGVKRDHFGALRFDCCGFTTYGVPVLCFGGFPPGTV 449
DB 837 aelprtmgyllhgcldvraglqghetraliktdcpvgtlqv---tpl---fwpv-spm 888
QY 450 FTQCLY 455
DB 889 wngcily 894


```
XX Sequence 333 AA:
SQ
Query Match 25.2%; Score 628.5; DB 22; Length 333;
Best Local Similarity 60.1%; Pred. No. 5.9e-47;
Matches 134; Conservative 15; Mismatches 55; Indels 19; Gaps 4;
QY 115 CKNVW---LMRKQILANKKEISKOOSTQEVYTWLKAFFSIREAHKSENLHPDNVTK 171
D 62 cpdlwnfelerddlgylveelskqsgisqntwtwllkafhfkretelkkslenlpdaae 121
QY 172 KKNPSESGFKIAAEICICNEELNVNPDNGENISWTCQSSQOSTIKSLAMPRR-KMPC 230
D 122 kkspfiaekfkpaaeiclsneepnv---ngvnvrtcgspqpfpsqalrhrkkwfh 177
QY 231 GTGPGSLCCVOPRDLVPCVPVNS-----AVASEGASPKPMQLDSGVPEVGAOKS 279
D 178 gqylgspccaaprelvpcvpatpmaerqgcrwamasegavpklwqlpuglepasaqks 237
QY 280 RIEWMBPPIRFQKIYGNPMPROKFAVGVSSWRTSARVYQKG 322
D 238 riewwppprfrqmygnacvprqkfaaevgpswrtssaaqpkq 280
RESULT 8
ABG16460
ID ABG16460 standard; Protein: 466 AA.
XX
AC ABG16460;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #16451.
XX
KM Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR N-PSDB; AAS80647.
XX
DR
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 46819; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
```

```
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 466 AA:
QY
Query Match 24.1%; Score 602; DB 22; Length 466;
Best Local Similarity 63.4%; Pred. No. 2.1e-44;
Matches 121; Conservative 14; Mismatches 32; Indels 24; Gaps 4;
QY 284 WEPPE-----IRFQKIYGNPMPROKFAVGVSSWRTSARVYQKGNGV 325
D 260 wpppdtmcskhgplrlhikrlfrfqkmygnawmpqrqkfavgsswrtssaramqkgnvg 319
QY 326 WEPPIHVPBGAPSSRAVRRSPSSRLQKGRSTSDLVHPEKSTDTQCCQVYKAAGMESVPY 385
D 320 weiphvpcgappsgavrrpssrrpqrstslncvppkaadtqgmkaagreaavpc 379
QY 386 KTVVAELRTVVGIVYLCHDIDVRRHGVKRDHFGALRFDCPTGRTYMGV-VPLCFGQGF 444
D 380 katgylelpkmetylhhgcdldvrhgvkghdfigalrfdcpagqtcmgvpvpl-----fwp 435
QY 445 FGTAVFTQCLY 455
D 436 i-splwngcily 445
RESULT 9
ABG00481
ID ABG00481 standard; Protein: 270 AA.
XX
AC ABG00481;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #472.
XX
KM Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR N-PSDB; AAS64668.
XX
DR
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 30840; 103pp; English.
```

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX Sequence 270 AA:
SQ

Query Match 23.0%; Score 573.5; DB 22; Length 270;
Best Local Similarity 48.3%; Pred. No. 3.1e-42;
Matches 128; Conservative 25; Mismatches 77; Indels 35; Gaps 5;

OY 202 GENIWTQGRSSQOSIKSLAMPFRKRWFCGNGPSGLCV-----QPRDLVPCVPVNS 253
DB 20 gddfyglaeelskqsglsxawe-----lgaicsshshgkrrgggraw----- 61
OY 254 AVASEGAPKPMQQLPSGVEPGAKSRLEWEPPIRFQKITGNPMPROKFAVGSSMR 313
DB 62 ataesaaskbgqgfhrstevpagaqsrlewepprltqrmynamwspkfaagsgpqr 121
OY 314 TSARVVGKNGWEPHARVPSGAPSSRAVRRSPSSRLQKGRSDSLQHVPEKSTDTQCQ 373
DB 122 tstrllrfgnvelnppnrvpcgklpsgavrtreppfrarnsgstnslnvcgtkdtlqrq 181
OY 374 PVKAGMSVPRKYVAELTKTVGTYLLHCHDLVDRHGVRKRDHFGALRFDCTGRTTWG 433
DB 182 pvkarrtealpckatgaelprktmgthlhcqcdlvtvgkghdfigaldfdcpgafqtcwg 241
OY 434 PVPLCFGQFPFGTAVER--OCLYL 456
DB 242 scs-----pfvlayfsylewyj 259

RESULT 10
ABG16421
ID ABG16421 standard; Protein: 439 AA.
XX
AC ABG16421:
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #16412.
XX
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
XX
DR N-PSDB: AAS80608.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 20; SEQ ID No 46780; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX Sequence 439 AA:
SQ

Query Match 22.9%; Score 570; DB 22; Length 439;
Best Local Similarity 62.2%; Pred. No. 1.2e-41;
Matches 117; Conservative 10; Mismatches 61; Indels 0; Gaps 0;

OY 257 SEGASPKPMQQLPSGVEPGAKSRLEWEPPIRFQKITGNPMPROKFAVGSSMRSTA 316
DB 233 slsvsagrpvsaaaagstalgksrlewepprtfqlklygnsvtpwqkfaagvgsptcsa 292
OY 317 RVVORGNGWEPHARVPSGAPSSRAVRRSPSSRLQKGRSDSLQHVPEKSTDTQCQPYK 376
DB 293 rlvwkgivgseppyrvtvgalpsgearrpqrngsrdsnlcyvpgkstdtcgpyk 352
OY 377 AAGMSVPRKYVAELTKTVGTYLLHCHDLVDRHGVRKRDHFGALRFDCTGRTTWGVPY 436
DB 353 aagregalpckatgaelstkgthlhcqcdlvtvgkghdfigaldfdcpgaftrltcmgpyt 412
OY 437 LCFGQFP 444
DB 413 plfwplsp 420

RESULT 11
ABG17512
ID ABG17512 standard; Protein: 439 AA.
XX
AC ABG17512:
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #17503.
XX
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;

Db 70 knpfiseekfkpaaetctsnrepnvkspqgwkmepwrt-tdlhgslsh-hrprglgen 126
OY 228 WFCGTGPGSLCCVOPRDLPVCPVNSAVASBEGASPKPMQLPSGVEPVGAKKSRIEWEPP 287
Db 127 gfvqgslgal-----hpnshnngfrgckpkpwlprdvsaasarkrleiewmkpp 175
OY 288 IRFOKITGNPMWMPROKFAVGVSSWRTSARVYOKGNVGMEPPHRYPSGAPSSRAVRRSP 347
Db 176 plfgkmygnvwmplqctfaagagsswrtatavqkgnvgwephrvtrtppsgalltrgpl 235
OY 348 SSRLOKGRSTDSLOHVEPKST 368
Db 236 sarpqngsstclshhvpngat 256

RESULT 13
ABG02632
ID ABG02632 standard; Protein: 261 AA.
XX
AC ABG02632;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #2623.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
DR N-PSDB: AAS66819.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 32991; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 261 AA;
Query Match 21.7%; Score 540.5; DB 22; Length 261;
Best Local Similarity 48.7%; Pred. No. 2.4e-39;
Matches 127; Conservative 15; Mismatches 68; Indels 51; Gaps 7;
OY 115 CKVNIW---LMRKOILAKKEISKOOSIOEVTWVLKAFSFIREAHRKSSENTLHPDNVTK 171
Db 40 cpdlwntelerddlgylveelstkgsglqe----- 69
OY 172 KKNPSEGRFKLAEDICONEELNV-NPDNGENISWTCORSSQOISLAMPRA---RK 227
Db 70 knpfiseekfkpaaetctsnrepnvkspqgwkmepwrt-tdlhgslsh-hrprglgen 126
OY 228 WFCGTGPGSLCCVOPRDLPVCPVNSAVASBEGASPKPMQLPSGVEPVGAKKSRIEWEPP 287
Db 127 gfvqgslgal-----hpnshnngfrgckpkpwlprdvsaasarkrleiewmkpp 175
OY 288 IRFOKITGNPMWMPROKFAVGVSSWRTSARVYOKGNVGMEPPHRYPSGAPSSRAVRRSP 347
Db 176 plfgkmygnvwmplqctfaagagsswrtatavqkgnvgwephrvtrtppsgalltrgpl 235
OY 348 SSRLOKGRSTDSLOHVEPKST 368
Db 236 sarpqngsstclshhvpngat 256

RESULT 14
ABG17044
ID ABG17044 standard; Protein: 766 AA.
XX
AC ABG17044;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #17035.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
DR N-PSDB: AAS81231.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 47403; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags

